

ALAGAPPA UNIVERSITY



[Accredited with 'A+' Grade by NAAC (CGPA:3.64) in the Third Cycle
and Graded as Category–I University by MHRD-UGC]

(A State University Established by the Government of Tamil Nadu)

KARAIKUDI – 630 003

Directorate of Distance Education

M.A. [Economics]

IV - Semester

362 44

COMPUTER APPLICATION IN ECONOMIC ANALYSIS

Authors:
Sanjay Saxena, Managing Director, Total Synergy Consulting Pvt. Ltd., New Delhi
Units: (1.0-1.5, 1.5.4, 1.5.7, 2, 3.3-3.4, 4.0-4.3)
Rajneesh Agrawal, Senior Scientist, Department of Information Technology, Government of India
Units: (4.4, 5.0-5.2, 5.4-5.8, 5.10-5.13)
C.B. Gupta, Former Director and Professor of Finance and Accounting at the Institute of Management Technology, Ghaziabad
Vijay Gupta, Retired Vice Chancellor of Sharda University, Greater Noida, Uttar Pradesh
Units: (9.3-9.5)
Pushpa Singh, Associate Professor, Delhi Technical Campus, Greater Noida
Narendra Singh, Lecturer, IT Sriram Group of Institutions, Greater Noida
Unit: (14)
Vikas Publishing House: Units: (1.5.1-1.5.3, 1.5.5-1.5.6, 1.6-1.10, 3.0-3.2, 3.5, 3.6-3.10, 4.5-4.10, 5.3, 5.9, 5.14-5.21, 6, 7, 8, 9.0-9.2, 9.6-9.12, 10, 11, 12, 13)



Vikas® is the registered trademark of Vikas® Publishing House Pvt. Ltd.

VIKAS® PUBLISHING HOUSE PVT. LTD.
E-28, Sector-8, Noida - 201301 (UP)
Phone: 0120-4078900 • Fax: 0120-4078999
Regd. Office: A-27, 2nd Floor, Mohan Co-operative Industrial Estate, New Delhi 1100 44
• Website: www.vikaspublishing.com • Email: helpline@vikaspublishing.com

Work Order No.AU/DDE/DE12-16/Printing of Course Materials /2020 dated 28.02.2020, Copies - 500

SYLLABI-BOOK MAPPING TABLE

Computer Application in Economic Analysis

Syllabi	Mapping in Book
BLOCK I: BASICS OF COMPUTERS **Unit-1:** Basics of Computer: Characteristics of Computer System, Evolution of Computers, Generations, Types of Computers. **Unit-2:** Basic Components of a Digital Computer: Control Unit, ALU, Input / Output Devices and Peripheral Devices. **Unit-3:** Computer Languages: Machine Language, Assembly Language and High Level Language. **Unit-4:** Operating System and NetWare: DOS, UNIX, Windows, LAN and WAN.	**Unit 1:** Basics of Computers **(Pages 1-24);** **Unit 2:** Basic Components of a Digital Computer **(Pages 25-73);** **Unit 3:** Computer Languages **(Pages 74-96);** **Unit 4:** Operating System and NetWare **(Pages 97-138)**
BLOCK II: BASIC CONCEPTS AND OPERATING SYSTEMS **Unit-5:** Basic Concepts of Internet: Modem, E-Mail, Websites, Address, Domain, Protocols, Types of Accounts, Search Engines, Browsing Web, Telnet, Usenet, AOL. **Unit-6:** Online Sources of Data: Books, Journals, Working Papers, Reports, Newspapers. **Unit-7:** Basics of MS-Office: MS-Word: The MS-Word Window, Entering, Selecting, Copying and Moving Text, Applying Fonts and Indenting Text, Creating Numbering and Bullets, Finding and Replacing Text, Spelling, Grammar and Thesaurus, Creating Page Headers and Footers, Constructing Tables and Mail Merge. **Unit-8:** MS-Excel Skills: Entering and Editing Cell Entries, Working with Numbers, Changing the Worksheet Layout, Formatting Text, Borders and Color, Printing in MS-Excel, Using Functions and References, Naming Ranges.	**Unit 5:** Basic Concepts of Internet **(Pages 139-200);** **Unit 6:** Online Sources of Data **(Pages 201-220);** **Unit 7:** Basics of MS-Office and MS-Word 2007 **(Pages 221-279);** **Unit 8:** MS Excel Skills **(Pages 280-338)**
BLOCK III: DATA ANALYSIS **Unit-9:** Data Analysis: Preparation of Frequency Distribution, Calculation of Mean, Standard Deviation, Coefficient of Variation, Correlation Coefficient, Regression Coefficients, Trend Line using Method of Least Squares and Moving Averages. **Unit-10:** SPSS: Basic Structure of an SPSS Data File, Using Data Editor, Reading Spread Sheet Data, Reading a Database, Reading Text Data. **Unit-11:** Data Definition: Preparation of Code Book, Data List, Begin Data, End Data, Defining Variable Type, Variable Tables, Value Labels, Missing Labels. **Unit-12:** Output Display: Interpretation of Output, Transfer of Data Display Formats, Creating and Editing Graphs, Tables and Diagrams, Bar, Line, Pie, Scatter, Histogram.	**Unit 9:** Data Analysis **(Pages 339-378);** **Unit 10:** Introduction to SPSS **(Pages 379-411);** **Unit 11:** Data Definition **(Pages 412-444);** **Unit 12:** Output Display **(Pages 445-502)**

BLOCK IV: COMPUTER APPLICATIONS TO ECONOMICS

Unit-13: Application to Economics: On-Line Banking; ATM's Electronic Stock Exchange; Electronic Trading; Data Sharing and Dissemination; Electronic Transaction; Document Delivery; Authentication and Validation Transaction Processing.

Unit-14: E-Business: Electronic Trading and Marketing; On-Line Shopping and Malls, B2B, B2C, Models, Document and Transaction Security and Digital Signature; Integrated Transaction on Mobile Platforms. E-Commerce Applications in India.

CONTENTS

INTRODUCTION

NOTES

Computers have brought about major changes in all spheres of life. Today it is extremely difficult to imagine the world without computers. Computers help us to communicate using modems, telephone and Wi-Fi facilities, and it seems as if you are sitting side by side and communicating directly with each other. This modern way of communicating has been made possible by computers. Through the Internet and e-mail, we now have the ability to communicate with anybody in any part of the world in a matter of minutes. The Internet links are computer networks across the world so that users can share resources and also communicate with each other. Some computers have direct access to all the facilities on the Internet and e-mail facilities of the Internet have been a boon to society, especially in terms of time saved. The study of computer networks and the Internet becomes essential to know more about computing techniques and communication technologies. Conveniences like ATM services, Internet, wireless telephony and electronic mail could not have been possible without computer networks.

Economics and finance have undergone radical changes during the Information Age, both in practice and in study. The ever-increasing computational and predictive capabilities of modern computers have made them invaluable tools for researching economic trends and developing strategies around them. Computational economics, a convergence of computer technology and economic theory, has changed the way companies and governments assess their economies.

The applications for computers in finance typically focus on investment planning and risk management. Using available statistical data, computers generate simulations that show the outcomes of investments under various situations and the potential for gains and losses. Using this information, companies develop plans to minimize potential losses and prepare for the different scenarios that may occur.

The emergence of e-commerce and online trading of goods, services and stocks has considerably changed the business methodologies and strategies. Many transactions, especially those between two businesses as opposed to a business and a consumer, are now performed online, with the exchange of information and digital purchases taking place instantly.

This book, *Computer Application in Economic Analysis*, is divided into four blocks that are further divided into fourteen units which will help you understand the basics of computer, basic components of a digital computer, computer languages, Operating System (OS) and NetWare (DOS, UNIX, Windows), LAN and WAN, basic concepts of Internet, e-mail, websites, search engines, online sources of data, basics of MS-Office, MS-Word window, constructing tables and mail merge, MS-Excel skills, entering and editing cell entries, changing the worksheet layout, formatting text, data analysis, frequency distribution, mean, standard deviation, coefficient of variation, correlation coefficient, regression coefficients, SPSS, basic

NOTES

structure of an SPSS data file, using data editor, reading spread sheet data, preparation of code book, output display, interpretation of output, creating and editing graphs, tables and diagrams, application to economics, on-line banking, ATM's, electronic trading, e-business, on-line shopping and malls, B2B and B2C models, document and transaction security and digital signature, and integrated transaction on mobile platforms.

The book follows the Self-Instruction Mode or the SIM format wherein each unit begins with an 'Introduction' to the topic followed by an outline of the 'Objectives'. The content is presented in a simple, organized and comprehensive form interspersed with 'Check Your Progress' questions and answers for better understanding of the topics covered. A list of 'Key Words' along with a 'Summary' and a set of 'Self Assessment Questions and Exercises' is provided at the end of the each unit for effective recapitulation. Logically arranged topics, relevant solved examples and illustrations have been included for better understanding of the topics.

BLOCK - I
BASICS OF COMPUTERS

NOTES

UNIT 1 BASICS OF COMPUTERS

Structure

1.0 INTRODUCTION

A computer is a machine that can be instructed to carry out sequences of arithmetic or logical operations automatically via computer programming. Modern computers have the ability to follow generalized sets of operations, called 'Programs'. These programs enable computers to perform an extremely wide range of tasks. A complete computer unit including the hardware, the operating system (main software), and peripheral equipment required and used for various operations/ tasks can be referred to as a computer system. This term may as well be used for a group of computers that are connected and work together, in particular a computer network or computer cluster. Any computer system essentially consists of three important parts, namely input devices, Central Processing Unit (CPU) and output devices. Computers are typically used as control systems for a wide variety of industrial and consumer devices, such as special purpose devices like microwave ovens and remote controls, factory devices, industrial robots, and mobile devices, such as smartphones. The Internet is run on computers and it connects hundreds of millions of other computers and their users.

NOTES

Fundamentally, a computer is a programmable machine. The two principal characteristics of a computer are that it responds to a specific set of instructions in a well-defined manner and it can execute a pre-recorded list of instructions (a program). Therefore, a computer may be defined as an electronic device that can operate upon information or data. The information is processed based on a set of instructions provided to generate the output. A computer is also capable of performing operations in numerical or logical terms. In computer technology, the term 'generation' is used to distinguish between varying hardware and software technologies. Computer software is a set of instructions containing the methodology to process raw data in a computer system. This set of instructions is also called a program. Hardware constitutes the electronic circuits that are used to build the computer. The electronic, magnetic and mechanical devices together are referred to as computer hardware. The various types of computers include desktop computers, notebook computers, laptop computers, personal computers (PCs), workstations, mainframe computers and supercomputers. The increasing popularity of the computer has proved that it is a powerful and useful tool. The size, shape, cost and performance of computers have changed over the years, but the basic logical structure has not.

In this unit, you will study about the basics of computer system, characteristics of computer system, evolution of computers, generations, and types of computers.

1.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the basics of a computer system. Define a computer
- Explain the evolution of computers
- Describe the characteristic advantages as well as disadvantages of computers
- Understand about the various generations of the computer's evolution
- Explain the classification of various types of computers

1.2 COMPUTER FUNDAMENTALS

Computers have undergone great transformation over the past decade; however, the basic logical structure remains the same. A computer primarily constitutes of three integral components, viz. input devices, Central Processing Unit (CPU) and output devices. The CPU constitutes of the main memory, the Arithmetic Logic Unit (ALU) and the Control Unit (CU).

Apart from these three basic components, computers have secondary storage devices known as auxiliary storage or backing storage that store data and instructions on a long-term basis.

NOTES

The following are the primary functions of a computer:

- **Inputting**: The process in which the user puts in a set of commands to process data into the computer system.
- **Storing**: The process of recording data and information so that it can be retrieved for use whenever required.
- **Processing**: This process implies performing arithmetic or logical operations on data to convert them into useful information. Arithmetic operations include addition, subtraction, multiplication and division, and logical operations include comparisons, such as equal to, less than and greater than, etc.
- **Outputting**: This is the process of providing results to the user. These can be in the form of visual display and/or printed reports.
- **Controlling**: This refers to directing the sequence and the manner in which all the previous functions are carried out.

A detailed description of the components that perform these tasks is as follows.

1. Input Unit

Programs and data are required to be present in a computer system before any operation can be performed. A program denotes the set of instructions which the computer has to carry out, and data is the information on which these instructions are to operate. If the task is to rearrange a list of telephone subscribers in alphabetical order, the sequence of instructions that will guide the computer through this operation is the program, while the list of names to be sorted is the data.

The Input Unit is responsible for transferring data and instructions from the external environment into the computer system. Instructions and data enter the input unit through the particular input device used (keyboard, scanner, card reader, etc.). These instructions and data are then converted into binary codes (computer-acceptable form) and sent to the computer system for further processing.

2. Central Processing Unit

The Central Processing Unit (CPU) is known as the brain of the computer. It is a blend of the control unit, the ALU and the primary memory that are described as follows:

- **Main Memory (Primary Storage):** The main memory or the primary storage of the computer system is responsible for storing all the instructions and data. The data is then transferred to the Arithmetic Logical Unit (ALU) for processing. After this, the final output is again stored back in the primary storage, until it is further sent to the output device.

  The primary storage also temporarily stores any intermediate result generated by the ALU. So data and instructions move frequently between the ALU

NOTES

and the primary storage before the processing is complete. It should be noted that no processing occurs within the primary storage.

- **Arithmetic Logic Unit:** In addition to the basic four arithmetic operations, viz. addition, subtraction, multiplication and division, the ALU also performs logic comparison operations including equal to, lesser than or greater than.
- **Control Unit:** The function of the control unit is to ensure that according to the stored instructions, the right operation is done on the right data at the right time. The control unit receives instructions and commands from the programs in the primary memory, processes them and ensures that the commands are executed in the desired order by all the other units of the computer system. In effect, the control unit is comparable to the central nervous system of the human body.

3. Output Unit

Computers understand, process data and return the output in a binary form. The basic function of the output unit is to convert these results into a human-readable form before providing the output through various output devices, such as terminals and printers.

The storage capacity of the primary memory of the computer is limited. Often, it is necessary to store large amounts of data. So, additional memory, called secondary storage or auxiliary memory, is used in most computer systems.

Secondary storage is storage other than the primary storage. These are peripheral devices connected to and controlled by the computer to allow permanent storage of data and programs. Usually, hardware devices like magnetic tapes and magnetic disks fall in this category.

Characteristics of Computers

The increasing popularity of the computer has proved that it is a powerful and useful tool. Its usefulness is due to its following advantages:

- **Speed**: Computers are very fast. They can process millions of instructions every second. The speed is related to the amount of data it processes and the time it takes to complete the processing task.
- **Storage**: Computers can store vast amounts of information in the form of files, which can be recalled at any time. These files help in easy and speedy retrieval of information. This type of storage is known as electronic storage system.
- **Accuracy**: In addition to being fast, computers are also accurate. The degree of accuracy for a particular computer depends upon its design. Most errors in computers are not of a technical nature and are human. Usually, programmers are responsible for these errors.

NOTES

- **Diligence**: Computers can perform any complicated task accurately without making any error. Computers do not suffer from carelessness, boredom or tiredness. Moreover, their efficiency does not decrease with age.
- **Versatility**: Computers perform various tasks depending upon the instructions given to them and their hardware characteristics. They are capable of performing any task, provided the task is reduced to a series of logical steps. A computer can be used to prepare a Word document and in between called to search for another document that is stored in its memory. It can perform both tasks simultaneously.
- **No IQ**: Computers do not have their own intelligence and their IQ (Intelligence Quotient) is zero. Hence, the user can and has to decide what tasks a computer should perform.
- **No Feelings**: Computers have no feelings because they are machines. They cannot make judgements as they process on the basis of a set of instructions, called programs, provided by the users.

Though computers can do better than human beings in terms of accuracy, speed and memory, there are certain disadvantages of computer systems as they depend on human beings for their operations and functions. The following are some of the disadvantages of computers:

- They depend on human beings who program them for efficient, accurate and fast functioning.
- Computers do not have their own intelligence and thus cannot think intelligently or work independently like human beings.
- They follow instructions given by programs or by users.
- They can neither take decisions nor can correct wrong instructions.
- Programmers or users maintain and update them.
- As with many other modern appliances, computers also need electric power to run.

Check Your Progress

1. What are the three important parts of a computer system?
2. Give three advantages of using computers.

1.3 EVOLUTION OF COMPUTERS

The first mechanical adding machine was invented by Blaise Pascal in 1642. Later, in 1671, Baron Gottfried Wilhelm von Leibniz of Germany invented the first calculator. Around this time, Herman Hollerith developed the concept of punched cards, which were extensively used as an input medium in mechanical adding machines.

NOTES

Charles Babbage, a 19th century professor at Cambridge University, is considered the father of the modern digital computer. During this period, mathematical and statistical tables were prepared by a group of clerks. However, utmost care and precaution could not eliminate human errors.

In 1842, Babbage came up with a new idea of the Analytical Engine, which was meant to be completely automatic. This machine was capable of performing basic arithmetic functions. However, these machines were difficult to manufacture because the precision required to manufacture them was not available at that time.

The following is a brief description of the various generations of computers.

- **Mark I Computer (1937–44)**: This was the first fully automatic calculating machine designed by Howard A. Aiken, the design of which was based on the technique of punching card machinery. In this technique, both mechanical and electronic components were used.
- **Atanasoff-Berry Computer (1939–42)**: This computer was developed by Dr John Atanasoff to solve certain mathematical equations. It used forty-five vacuum tubes for internal logic and capacitors for storage.
- **ENIAC (1943–46)**: The Electronic Numerical Integrator and Computer (ENIAC) was the first electronic computer developed for military requirements and was used for many years to solve ballistic problems.
- **EDVAC (1946–52)**: One of the drawbacks of ENIAC was that its programs were wired on boards, which made it difficult to change them. To overcome the drawbacks of ENIAC, the Electronic Discrete Variable Automatic Computer (EDVAC) was designed. The basic idea behind this concept was that sequences of instructions could be stored in the memory of the computer for automatically directing the flow of operations.
- **EDSAC (1947–49)**: Professor Maurice Wilkes developed the Electronic Delay Storage Automatic Calculator (EDSAC), by which addition and multiplication operations could be accomplished.
- **UNIVAC I (1951)**: The UNIVersal Automatic Computer (UNIVAC) was the first digital computer to be installed in the Census Bureau in 1951 and was used continuously for 10 years. In 1952, International Business Machines (IBM) introduced the 701 commercial computers. These computers could be used for scientific and business purposes.

Check Your Progress

3. When was the first mechanical adding machine invented and by whom?
4. Who is considered as the father of modern digital computer?

1.4 GENERATIONS OF COMPUTER

The history of computer development can be divided into different phases which are often referred to as generations of computing devices. 'Generation' in computer terminology is a 'step' in technology. Each generation of computers is characterized by a major technological development that fundamentally changes the way computers operate, resulting in increasingly smaller, cheaper, more powerful, efficient and reliable devices with decreasing energy consumption and lesser generation of heat.

Originally, the term 'generation' was used to distinguish between varying hardware technologies, but nowadays, it includes both hardware and software.

The following are the characteristics of each generation of computers:

1. First-Generation (1940–1956): Vacuum Tubes

The first computers used vacuum tubes in their electronic circuits and magnetic drums for memory. A vacuum tube was a delicate glass device that used filaments as a source of electrons and could control and amplify electronic signals. Figure 1.1 displays a vacuum tube.



Fig. 1.1 *A Vacuum Tube*

These computers could perform computations in milliseconds but were enormous in size, occupying almost an entire room. They were very expensive to operate and in addition to using a great deal of electricity, generated a lot of heat resulting in malfunctioning.

First-generation computers relied on machine language (binary-coded programs) to perform operations and could solve only one problem at a time. Input was based on punch cards and paper tape, and output was displayed on printouts.

Early computers like ENIAC, EDVAC, UNIVAC I all can be classified as first-generation computers.

2. Second-Generation (1956–1963): Transistors

Transistors developed in 1947 replaced vacuum tubes in the second-generation computers. The transistor was far superior compared to vacuum tube, making computers smaller, faster, cheaper, more energy-efficient and more reliable than their first-generation predecessors. Although transistors also generated a great deal of heat that could damage the computer, it was a great improvement over the

NOTES

vacuum tube. Second-generation computers still relied on punched cards for input and printouts for output.

The cryptic binary machine language was followed by the symbolic or assembly language that allowed programmers to specify instructions in words. High-level programming languages like COBOL and FORTRAN were also being developed at this time.

These were also the first computers that stored their instructions in the memory, which advanced from magnetic drum to magnetic core technology. The first computers of this generation were specifically developed for the atomic energy industry.

3. Third-Generation (1964–1971): Integrated Circuits

Transistors were clearly an improvement over the vacuum tube but still generated a lot of heat resulting in computer damage.

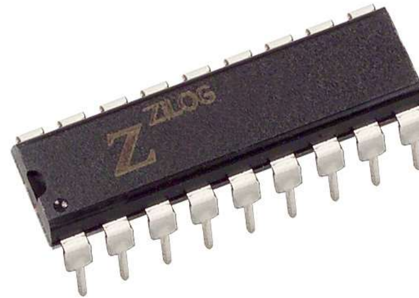


Fig. 1.2 *An IC Chip*

The development of integrated circuit (see Figure 1.2) by Jack Kilby in 1958, an engineer with Texas Instruments, was the greatest achievement of the third-generation of computers.

Instead of punched cards and printouts, users interacted with third-generation computers through devices like keyboards and monitors. They also interfaced with an operating system that allowed the device to run many different applications at one time with a central program that monitored the memory.

Now, the computers became accessible to the masses because they were substantially smaller and cheaper than their predecessors.

4. Fourth-Generation (1971–Present): Microprocessors

Large Scale Integration (LSI) were developed which could fit hundreds of components onto a single chip. By 1980s, Very Large Scale Integration (VLSI) squeezed thousands of components onto a single chip. Ultra Large Scale Integration (ULSI) increased that number to millions.

NOTES

The ability to fit so much processing capability in an area so small, helped to reduce the size and price of the computers. It also increased its power, efficiency and reliability.

Initially, the IC technology was used only for constructing the processor, but it was soon discovered that the same technology could also be used for the construction of memory. The first memory chip was constructed in 1970 and could hold 256 bits. Figure 1.2 displays an IC chip.

As more and more components were fabricated on a single chip, fewer and fewer chips were needed to construct the processor. The Intel 4004 chip, developed in 1971, located all the components of the computer — from Central Processing Unit and Memory to Input/Output controls — on a single chip. This was the first microprocessor. Figure 1.3 displays the intel pentium microprocessor chip.



Fig. 1.3 *The Intel Pentium Microprocessor Chip*

IBM introduced its first computer in 1981 for the home users, and in 1984 Apple introduced the Macintosh. Microprocessors also advanced from the realm of desktop computers to advanced technologies and many areas of life as more and more everyday devices began to use microprocessors.

As computers increased in computing power, it was possible to connect them together to form networks, which eventually led to the development of the Internet. Fourth-generation computers also marked the development of GUIs, the mouse and various handheld devices.

5. Fifth-Generation (Present and Beyond): Artificial Intelligence

The fifth-generation computers are being developed using the technology of artificial intelligence; for instance, voice recognition systems. Parallel processing and supercomputers have lead to the further development of artificial intelligence. In the future, quantum computation and molecular technology will tremendously transform computers. The fifth-generation aims at creating devices that respond to input in natural language and are capable of learning and self organization.
Table 1.1 provides a list of various computer generations.

NOTES

Table 1.1 *Generation of Computers*

Generation	Time	Hardware	Software	Features	Examples
I	**1942-1955**	**Vacuum Tubes**	Machine Language (Binary Language)	High-speed electronic switching device; memory type was electromagnetic; bulky in size; generated a large amount of heat; frequent technical faults; required constant maintenance; used for scientific purposes; air-conditioning required	ENIAC, EDVAC, EDSAC, UNIVAC I
II	**1955-1964**	**Transistors**	High-level languages FORTRAN, COBOL, ALGOL, SNOBOL	Better electronic switching devices than vacuum tubes; made of germanium semiconductors; memory type was magnetic cores; powerful and more reliable; easy to handle; much smaller than vacuum tubes; generated less heat as compared to vacuum tubes; used for business and industries for commercial data processing; air-conditioning required	Livermore Atomic Research Computer (LARC), IBM
III	**1964-1975**	**Integrated Circuits (ICs)** made up of transistors, resistors and capacitors fixed on single silicon chip	High-level languages PL/1, PASCAL, BASIC, VISUAL BASIC, C, C++, C#, Java	ICs were smaller than transistors; consumed less power; dissipated less heat as compared to transistors; more reliable and faster than earlier generations; capable of performing about 1 million instructions per second; large storage capacity; used for both scientific and commercial purposes; air-conditioning required	Mainframe, Minicomputers
IV	**1975-1989**	**Microprocessor** made up of Large Scale Integration Circuits (LSI) and Very Large Scale Integration Circuits (VLSI)	Advanced Java (J2EE, JDO, JavaBeans), PHP, HTML, XML, SQL	Microprocessor had control on logical instructions and memory; semiconductor memories; personal computers were assembled; used in LAN and WAN to connect multiple computers at a time; used graphical user interface; smaller, more reliable and cheaper than third-generation computers; larger primary and secondary storage memories; had Computer Supported Cooperative Working (CSCW); air-conditioning not required	Personal Computers (PCs), LAN, WAN, CSCW
V	**1989-Present**	Ultra Scale Large Integration (USLI), Optical Disks	Artificial Intelligence, PROLOG, OPS5, Mercury	PCs were assembled – portable and non-portable, powerful desktop PCs and workstations; less prone to hardware failure; user-friendly features – Internet, e-mailing; air-conditioning not required	Portable PCs, Palmtop Computers, Laptop

Check Your Progress

5. What development was seen in the fourth generation microcomputers?
6. Give examples of third generation computers.

1.5 CLASSIFICATION OF TYPES OF COMPUTERS

Computers can be classified on the basis of their size, processing speed and cost. The various types of computers are:

- Personal Computers (PCs)
- Workstations
- Notebook/Laptop Computers

- Tablet PC
- PDA
- Mainframe Computers
- Supercomputers

NOTES

1.5.1 Analog Computers

These types of computers are involved in industrial process controls and measure physical quantities, like pressure, temperature, etc. These computers do not use binary digits but use electrical signals to provide output with electrical resistance, voltage, etc. The memory of these computers is not much and they can be used only for specific calculations but their speed is more than digital computers.

These electrical properties allow calculations to be performed in real time or even faster at the speed of light. The main mathematical operations it applies include summation, inversion, exponentiation, logarithm, integration, differentiation, multiplication and division.

1.5.2 Digital Computers

These types of computers are primarily involved in data processing and problem-solving for specific programs. In digital computers, data is stored as digits (numbers) and processes. Letters, words, symbols and complete texts are digitally represented, i.e., using only two digits '0' and '1'. Digital computers have a lot of memory for storing data.

Digital computers constitute input-output devices, main memory, control unit and arithmetic logic unit. Data is processed with logical circuits, also known as digital circuits. All the circuits processing data inside a computer function in an extremely synchronized mode; which is further controlled using a steady oscillator acting as the computer's 'clock'. Hence, the digital computers operate on very high speed and are able to perform trillions of logical or arithmetic operations per second to provide quick solution to problems, which is not possible for a human being to do manually.

1.5.3 Hybrid Computers

Hybrid computers are a mixture of digital and analog computers. A hybrid computer uses the best characteristics of digital and analog computers. It helps the user to process both continuous and discrete data. Hybrid computers are generally used for weather forecasting and industrial process control.

The digital component basically functions as a controller to provide logical operations, whereas the analog component provides solutions of differential equations. Remember that the hybrid computers are different from hybrid systems. The hybrid system is a digital computer equipped with an analog-to-digital converter for input and a digital-to-analog converter for output. The term 'hybrid computer' represents a combination of different digital technologies to process specific applications with the help of various specific processor technologies.

NOTES

1.5.4 General Purpose Computers

Workstations are high-end, general-purpose computers designed to meet the computing needs of engineers, architects and other professionals who need computers with greater processing power, larger storage and better graphic display facilities. These are commonly used for Computer-Aided Design (CAD) and for multimedia applications, such as creating special audio-visual effects for television programmes and movies. A workstation looks like a PC and can be used by only one person at a time. The characteristics of a workstation, which are often used to differentiate it from a PC, are as follows:

- **Display Facility**: Most workstations have a large-screen monitor (21 inches or more) capable of displaying high-resolution graphics as compared to PCs, which have a small-screen monitor (19 inches or less).
- **Storage Capacity**: Workstations have a larger main memory than PCs, which have only a few hundred MB of main memory. The hard disk capacity of workstations is also more than that of PCs.
- **Processing Power**: The processing power of workstations is several times greater than that of PCs.
- **Operating System**: PCs can run any of the five major operating systems—MS-DOS, MS-Windows, Windows-NT, Linux and UNIX—but all workstations generally run the Unix operating system or a variation of it, such as AIX (used in IBM workstations), Solaris (used in SUN workstations) and HPUX (used in HP workstations).
- **Processor Design**: PCs normally use CPUs based on the Complex Instruction Set Computer (CISC) technology, whereas workstation CPUs are based on the Reduced Instruction Set Computer (RISC) technology.

1.5.5 Special Purpose Computer

A special purpose computer is a digital or an analog computer specifically designed to perform desired specific task. These are high-performance computing systems with special hardware architecture, which is dedicated to solve a specific problem. This is performed with the help of specially programmed FPGA chips or custom VLSI chips. They are used for special applications, for example, astrophysics computations, GRAPE-6 (for astrophysics and molecular dynamics), Hydra (for playing chess), MDGRAPE-3 (for protein structure computations), etc.

1.5.6 Micro, Mini, Mainframe and Supercomputers

These are as follows:

(i) Microcomputers

Microcomputers are developed from advanced computer technology. They are commonly used at home, classroom and in the workplace. Microcomputers are

NOTES

called home computers, personal computers, laptops, personal digital assistants, etc. They are powerful and easy to operate. In recent years, computers were made portable and affordable. The major characteristics of a microcomputer are as follows:

- Microcomputers are capable of performing data processing jobs and solving numerical programs. Microcomputers work rapidly like minicomputers.
- Microcomputers have reasonable memory capacity which can be measured in megabytes.
- Microcomputers are reasonably priced. Varieties of microcomputers are available in the market which can be as per the requirement of smaller business companies and educational institutions.
- Processing speed of microcomputers is measured in megahertz. A microcomputer running at 90MHz works approximately at 90 MIPS.
- Microcomputers have drives for floppy disk, compact disk and hard disks.
- Only one user can operate a microcomputer at a time.
- Microcomputers are usually dedicated to one job. Millions of people use microcomputers to increase their personal productivity.
- Useful accessory tools, such as clock, calendar, calculator, daily schedule reminders, scratch pads, etc., are available in a microcomputer.
- Laptop computers, also called notebook computers, are microcomputers. They use the battery power source. Laptop computers have a keyboard, mouse, floppy disc drive, CD drive, hard disk drive and monitor. Laptop computers are expensive in comparison to personal computers.

(ii) Minicomputers

Minicomputers are a cheaper version of mainframe computers. The processing power and cost of a minicomputer are less than that of the mainframe. The minicomputers have big memory sizes and faster processing speed compared to the microcomputer. Minicomputers are also called workgroup systems because they are well suited to the requirements of the minor workgroups within an organization. The major characteristics of a minicomputer are as follows:

- Minicomputers have great problem solving capabilities.
- Minicomputers have reasonable memory capacity which can be measured in megabytes or gigabytes.
- Minicomputers have quick processing speeds and operating systems facilitated with multitasking and network capabilities.
- Minicomputers have drives for floppy disk, magnetic tape, compact disk, hard disks, etc.
- Minicomputers can serve as network servers.
- Minicomputers are used as a substitute of one mainframe by big organizations.

NOTES

(iii) Mainframe Computers

Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it. The user terminals act as access stations and may be located in the same building (refer to Figure 1.4).



Fig. 1.4 *Mainframe Computer*

Mainframe computers are bigger and more expensive than workstations. They look like a row of large file cabinets and need a large room with closely monitored humidity and temperature levels. A mainframe system of lower configuration is often referred to as a minicomputer system. The various components of a mainframe computer are as follows:

- **Host, Front-End and Back-End Computers**: A mainframe system consists of several computers, such as a host computer that carries out most of the computations and has direct control over all other computers. The front-end portion is used for handling communications to and from all the user terminals connected to the mainframe computer. The back-end portion is used to handle data input/output operations. The host computer and other computers are located in the systems room, to which entry is restricted to system administrators and maintenance staff only.
- **Consoles**: Console terminals are directly connected to the host computer and are mainly used by the system administrator to perform certain administrative tasks like installing new software on the system, taking system backups and changing the configuration of the system.
- **Storage Devices**: A mainframe computer has several magnetic disk drives directly connected to the back-end computer. The host computer, via the back-end computer gets all data from these magnetic disks. In addition, a mainframe computer also has a few tape drives and a magnetic

tape library (located in the systems room) for restoration and backup of data. The tape drives are present in the users' room, so that users' tapes can be used for input and output.

- **User Terminals**: User terminals are used to access the required stations, which may be present at different locations. Since mainframe computers support multiprogramming with time-sharing, they can run different operating systems for multiple users at the same time.
- **Output Devices**: A mainframe computer has several output devices like printers and plotters, connected to the back-end computer, so that these devices can be used for taking outputs by the users.

NOTES

(iv) Supercomputers

Supercomputers are the most powerful and expensive computers available today. They are mainly used for processing scientific applications that involve tasks with highly complex calculations and for solving problems with mechanical physics, such as weather forecasting and climate research systems, nuclear weapon simulation and simulation of automated aircrafts. Supercomputers are mainly used by military organizations, major research and development centres, universities and chemical laboratories.

Supercomputers use multiprocessing and parallel processing technologies to solve complex problems quickly. They use multiprocessors, which help the user to divide a complex problem into smaller problems. A parallel program is written in a manner that can break up the original problem into smaller modules. Supercomputers also support multiprogramming, which allows simultaneous access to the computer by multiple users. Some of the manufacturers of supercomputers are IBM, Silicon Graphics, Fujitsu and Intel.

1.5.7 Personal Computers (PCs)

A PC or Personal Computer is a small single-user microprocessor-based computer that sits on your desktop and is generally used at homes, offices, and schools. As the name implies, PCs were mainly designed to meet the personal computing needs of individuals. Personal computers are used for preparing normal text documents, spreadsheets with predefined calculations and business analysis charts, database management systems, accounting systems and also for designing office stationery, banners, bills and handouts.

The configuration varies from one PC to another depending on its usage. However, it consists of a CPU or system unit, a monitor, a keyboard and a mouse. It has a main circuit board or motherboard (consisting of the CPU and the memory), hard disk storage, floppy disk drive, CD-ROM drive and some special add-on cards (like Network Interface Card) and ports for connecting peripheral devices like printers.

PCs are available in two models—desktop and tower. In the desktop model, the monitor is positioned on top of the system unit, whereas in the tower model the

system unit is designed to stand by the side of the monitor or even on the floor to save desktop space. Due to this feature, the tower model is more popular.

Some popular operating systems for PCs are MS-DOS, MS-Windows, Windows-NT, Linux and Unix. Most of these operating systems can perform many functions at the same time which eases operation and saves time when a user has to switch between two or more applications while performing a job. Some leading PC manufacturers are IBM, Apple, Compaq, Dell, Toshiba and Siemens.

Types of Personal Computers

Different types of personal computers are as follows:

1. Notebook/Laptop Computers

Notebook computers are battery-operated personal computers. Smaller than the size of a briefcase, these are portable computers and can be used in places like libraries, in meetings or even while travelling. Popularly known as laptop computers, or simply laptops. Notebook computers or laptops are usually more expensive as compared to desktop computers though they have almost the same functions, but since they are sleeker and portable they have a complex design and are more difficult to manufacture. The Laptop computers have large storage space and other peripherals, such as serial port, PC card, modem or network interface card, CD-ROM drive and printer. They can also be connected to a network to download data from other computers or to the Internet. A notebook computer has a keyboard, a flat screen with Liquid Crystal Colour (LCD) Display (see Figure 1.5), and can also have a trackball and a pointing stick.



Fig. 1.5 *A Laptop Computer*

A notebook computer uses the MS-DOS or WINDOWS operating system. The data processing capability of a notebook computer is as good as an ordinary PC because both use the same type of processor, such as an Intel Pentium processor. However, a notebook computer generally has lesser hard disk storage than a PC.

2. Tablet PC

Tablet PC is a mobile computer that looks like a notebook or a small writing slate but uses a stylus pen or your finger tip to write on the touch screen. It saves whatever you scribble on the screen with the pen, in the same way as you have written it. The same picture can then be converted to text with the help of a HR (Hand Recognition) software.

NOTES

3. PDA

A Personal Digital Assistant (PDA) is a small, palm sized, hand-held computer which has a small color touch screen with audio and video features. They are nowadays used as smart phones, web enabled palmtop computers, portable media players or gaming devices.

Most PDAs today typically have a touch screen for data entry, a data storage/ memory card, Bluetooth, Wi-Fi or an infrared connectivity and can be used to access the Internet and other networks.

Check Your Progress

7. Name the various types of computers.
8. Describe the characteristics of minicomputers.
9. What is a tablet PC?

1.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A computer primarily constitutes three integral components, viz. input device, central processing unit (CPU) and output device.
2. The three advantages of using computers are:
   - Speed: Computers are very fast. They can process millions of instructions every second. The speed is related to the amount of data it processes and the time it takes to complete the processing task.
   - Storage: Computers can store vast information in the form of files, which can be recalled at any time. These files help in easy and speedy retrieval of information. This type of storage is known as electronic storage system.
   - Accuracy: In addition to being fast, computers are also accurate. The degree of accuracy for a particular computer depends upon its design. Most errors in computers are not of a technical nature and are human. Usually, programmers are responsible for these errors.
3. The first mechanical adding machine was invented by Blaise Pascal in 1642.
4. Charles Babbage, a 19th century professor at Cambridge University, is considered the father of the modern digital computer.

NOTES

5. In the fourth generation computers, Large Scale Integration (LSI) were developed which could fit hundreds of components onto a single chip. By 1980s, Very Large Scale Integration (VLSI) squeezed thousands of components onto a single chip and later. Ultra Large Scale Integration (ULSI) increased that number to millions.
6. Mainframes as well as minicomputers are some of the examples of third generation computers.
7. The various types of computers are:
   - Personal Computers
   - Workstations
   - Notebook/Laptop Computers
   - Tablet PC
   - PDA
   - Mainframe Computers
   - Supercomputers
8. The major characteristics of a minicomputer are as follows:
   - Minicomputers have great problem solving capabilities.
   - These have reasonable memory capacity which can be measured in megabytes or gigabytes.
   - These types of computers have quick processing speeds and operating systems facilitated with multitasking and network capabilities.
   - These have drives for floppy disk, magnetic tape, compact disk, hard disks, etc.
   - These can serve as network servers.
   - Minicomputers are used as a substitute of one mainframe by big organizations.
9. Tablet PC is a mobile computer that looks like a notebook or a small writing slate but uses a stylus pen or your finger tip to write on the touch screen. It saves whatever you scribble on the screen with the pen, in the same way as you have written it.

1.7 SUMMARY

- Computers have undergone great transformation over the past decade; however, the basic logical structure remains the same.
- A computer primarily constitutes of three integral components, viz. input devices, Central Processing Unit (CPU) and output devices. The CPU constitutes of the main memory, the Arithmetic Logic Unit (ALU) and the Control Unit (CU).

NOTES

- Storing refers to the process of recording data and information so that it can be retrieved for use whenever required.
- Processing implies performing arithmetic or logical operations on data to convert them into useful information.
- Arithmetic operations include addition, subtraction, multiplication and division, and logical operations include comparisons, such as equal to, less than and greater than, etc.
- Programs and data are required to be present in a computer system before any operation can be performed.
- A program denotes the set of instructions which the computer has to carry out, and data is the information on which these instructions are to operate.
- The Input Unit is responsible for transferring data and instructions from the external environment into the computer system. Instructions and data enter the input unit through the particular input device used (keyboard, scanner, card reader, etc.). These instructions and data are then converted into binary codes (computer-acceptable form) and sent to the computer system for further processing.
- The Central Processing Unit (CPU) is known as the brain of the computer. It is a blend of the control unit, the ALU and the primary memory.
- The main memory or the primary storage of the computer system is responsible for storing all the instructions and data. The data is then transferred to the Arithmetic Logical Unit (ALU) for processing. After this, the final output is again stored back in the primary storage, until it is further sent to the output device.
- The function of the control unit is to ensure that according to the stored instructions, the right operation is done on the right data at the right time. The control unit receives instructions and commands from the programs in the primary memory, processes them and ensures that the commands are executed in the desired order by all the other units of the computer system.
- The basic function of the output unit is to convert the processed results into a human-readable form before providing the output through various output devices, such as terminals and printers.
- The storage capacity of the primary memory of the computer is limited. Often, it is necessary to store large amounts of data. So, additional memory, called secondary storage or auxiliary memory, is used in most computer systems.
- Computers are very fast. They can process millions of instructions every second. The speed is related to the amount of data it processes and the time it takes to complete the processing task.

NOTES

- Charles Babbage, a 19th century professor at Cambridge University, is considered the father of the modern digital computer. During this period, mathematical and statistical tables were prepared by a group of clerks. However, utmost care and precaution could not eliminate human errors.
- The history of computer development can be divided into different phases which are often referred to as generations of computing devices.
- ‘Generation’ in computer terminology is a ‘step’ in technology. Each generation of computers is characterized by a major technological development that fundamentally changes the way computers operate, resulting in increasingly smaller, cheaper, more powerful, efficient and reliable devices with decreasing energy consumption and lesser generation of heat.
- Computers can be classified on the basis of their size, processing speed and cost. The various types of computers are Personal Computers (PCs), Workstations, Notebook/Laptop Computers, Tablet PC, PDA, Mainframe Computers and Supercomputers.
- Analog types of computers are involved in industrial process controls and measure physical quantities, like pressure, temperature, etc. These computers do not use binary digits but use electrical signals to provide output with electrical resistance, voltage, etc.
- Digital types of computers are primarily involved in data processing and problem-solving for specific programs. In digital computers, data is stored as digits (numbers) and processes. Letters, words, symbols and complete texts are digitally represented, i.e., using only two digits ‘0’ and ‘1’. Digital computers have a lot of memory for storing data.
- Hybrid computers are a mixture of digital and analog computers. A hybrid computer uses the best characteristics of digital and analog computers. It helps the user to process both continuous and discrete data. Hybrid computers are generally used for weather forecasting and industrial process control.
- Workstations are high-end, general-purpose computers designed to meet the computing needs of engineers, architects and other professionals who need computers with greater processing power, larger storage and better graphic display facilities. These are commonly used for Computer-Aided Design (CAD) and for multimedia applications, such as creating special audio-visual effects for television programmes and movies.
- A special purpose computer is a digital or an analog computer specifically designed to perform desired specific task. These are high-performance computing systems with special hardware architecture, which is dedicated to solve a specific problem. This is performed with the help of specially programmed FPGA chips or custom VLSI chips.

NOTES

- Microcomputers are developed from advanced computer technology. They are commonly used at home, classroom and in the workplace. Microcomputers are called home computers, personal computers, laptops, personal digital assistants, etc. They are powerful and easy to operate.
- Minicomputers are a cheaper version of mainframe computers. The processing power and cost of a minicomputer are less than that of the mainframe. The minicomputers have big memory sizes and faster processing speed compared to the microcomputer.
- Minicomputers are also called workgroup systems because they are well suited to the requirements of the minor workgroups within an organization.

- Mainframe computers are generally used for handling the needs of information processing of organizations like banks, insurance companies, hospitals and railways. This type of system is placed in a central location with several user terminals connected to it. The user terminals act as access stations and may be located in the same building.
- Supercomputers are the most powerful and expensive computers available today. They are mainly used for processing scientific applications that involve tasks with highly complex calculations and for solving problems with mechanical physics, such as weather forecasting and climate research systems, nuclear weapon simulation and simulation of automated aircrafts.
- Personal Computer (PC) is a small single-user microprocessor-based computer that sits on your desktop and is generally used at homes, offices, and schools. Personal computers are used for preparing normal text documents, spreadsheets with predefined calculations and business analysis charts, database management systems, accounting systems and also for designing office stationery, banners, bills and handouts.
- Notebook computers are battery-operated personal computers. Smaller than the size of a briefcase, these are portable computers and can be used in places like libraries, in meetings or even while travelling. Popularly known as laptop computers, or simply laptops.
- Notebook computers or Laptops are usually more expensive as compared to desktop computers though they have almost the same functions, but since they are sleeker and portable they have a complex design and are more difficult to manufacture.
- The Laptop computers have large storage space and other peripherals, such as serial port, PC card, modem or network interface card, CD-ROM drive and printer.
- Tablet PC is a mobile computer that looks like a notebook or a small writing slate but uses a stylus pen or your finger tip to write on the touch screen. It saves whatever you scribble on the screen with the pen, in the same way as

you have written it. The same picture can then be converted to text with the help of a HR (Hand Recognition) software.

- A Personal Digital Assistant (PDA) is a small, palm sized, hand-held computer which has a small colour touch screen with audio and video features. They are nowadays used as smart phones, web enabled palmtop computers, portable media players or gaming devices.
- Most PDAs today typically have a touch screen for data entry, a data storage/ memory card, Bluetooth, Wi-Fi or an infrared connectivity and can be used to access the Internet and other networks.

1.8 KEY WORDS

- **Inputting:** It is the process wherein the user feeds in the set of commands or instructions to process data into the computer system.
- **Storing:** The process of recording data and information so that it can be retrieved for use whenever required.
- **Processing:** Performing arithmetic or logical operations on data to convert them into useful information.
- **Outputting:** The process of providing results to the user which in the form of visual display and/or printed reports.
- **Controlling:** Directing the sequence and the manner in which all these previous tasks are carried out.
- **Hybrid computers:** Hybrid computers are the combination of digital and analog computers and uses the best features of both.
- **Tablet PC:** Tablet PC is a mobile computer that looks like a notebook or a small writing slate but uses a stylus pen or your finger tip to write on the touch screen.
- **PDA:** A Personal Digital Assistant (PDA) is a small, palm sized, hand-held computer which has a small color touch screen with audio and video features.

1.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a computer?
2. Define the advantages and disadvantages of computers.
3. Name the stages of evolution of computers.
4. What is ‘Generation’ in computers?

5. Differentiate between digital and analog computers.
6. What are supercomputers?
7. What are the key features of general purpose computers?
8. What is special purpose computer?
9. Why are Personal Computers (PCs) becoming more popular?

Long-Answer Questions

1. Briefly discuss the significance of computer technology.
2. Describe the basic tasks that can be performed by a computer system.
3. Explain the various stages of evolution of computers giving examples of each type.
4. Briefly describe the various generations of the computer giving examples of each type.
5. Explain in detail the classification system of types of computers giving examples of each type.
6. Briefly explain the features of Micro, Mini, Mainframe and Supercomputers.
7. Write short notes on the following:
    (a) Analog Computers
    (b) Digital Computers
    (c) Hybrid Computers
    (d) General Purpose Computers
    (e) Special Purpose Computer
    (f) Personal Computers

1.10 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

NOTES

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 2 BASIC COMPONENTS OF A DIGITAL COMPUTER

NOTES

Structure

2.0 INTRODUCTION

A computer is a combination of hardware and software resources which integrate together and provides various functionalities to the user. Hardware are the physical components of a computer, such as the processor, memory devices, monitor, keyboard, etc., while software is the set of programs or instructions that are required by the hardware resources to function properly. The basic components that help in the working of a computer, i.e., the Input- Process- Output, are termed as the functional components of a computer. It needs certain input, processes that input and produces the desired output. The input unit takes the input, the central processing unit does the processing of data and the output unit produces the output. The memory unit holds the data and instructions during the processing. A digital computer can be defined as a programmable machine which reads the binary data passed as instructions, processes this binary data, and displays a calculated digital output in the user readable format. Therefore, the digital computers work on the digital data to perform various computational tasks. Digital computers use the binary number system, which has two digits, '0' and '1'. A binary digit is called a bit. Information is represented in digital computers in groups of bits. By using various coding techniques, groups of bits can be made to represent not only binary numbers but also other discrete symbols, such as decimal digits or letters of the alphabet.

The input unit consists of input devices that are attached to the computer. These devices take input and convert it into binary language that the computer understands. Some of the common input devices are keyboard, mouse, joystick, scanner, etc. The processor of the Central Processing Unit (CPU) processes the information which is entered into the computer by the input device. The CPU is called the brain of the computer because it is the control centre of the computer. The CPU has three main components which are responsible for different functions, namely the Arithmetic Logic Unit (ALU), Control Unit (CU) and Memory registers. The ALU, as its name suggests performs mathematical calculations and takes logical decisions. Arithmetic calculations include addition, subtraction, multiplication and division. The Control unit coordinates and controls the data flow in and out of CPU and also controls all the operations of ALU, memory registers and also input/output units. It is also responsible for carrying out all the instructions stored in the program.

In computing, Input / Output or I/O devices refers to the communication between an information processing system, such as a computer, and the outside world, possibly a human or another information processing system. Inputs are the signals or data received by the system and outputs are the signals or data sent from it. I/O devices are the pieces of hardware used by a human (or other system) to communicate with a computer. For instance, a keyboard or computer mouse is an input device for a computer, while monitors and printers are output devices.

In this unit, you will study about the basic components of a digital computer, Control Unit (CU), Arithmetic Logic Unit (ALU), Input / Output (I/O) devices and peripheral devices.

2.1 OBJECTIVES

After going through this unit, you will be able to:

- Explain the various components of digital computers
- Explain the use of different input devices
- Describe various output devices with their uses
- Understand about the various storage devices and their features

2.2 COMPONENTS OF A DIGITAL COMPUTER

In order to transfer data to the memory of the computer input devices are used. The Arithmetic Logic Unit (ALU) is responsible for calculations, to which this data from the memory is stored. Once the calculations are done, the data is transferred back to the memory. The memory is responsible for storing data, according to which different functions are carried out. This memory is also known as the main memory or the Immediate Access Store (IAS).

The control unit is responsible for controlling various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer. It is the main function of the control unit to make sure that the instructions are correctly followed and all operations are done exactly according to the correct instructions at the correct time. This process leads to outcomes that are stored in memory. Figure 2.1 displays a computer system.

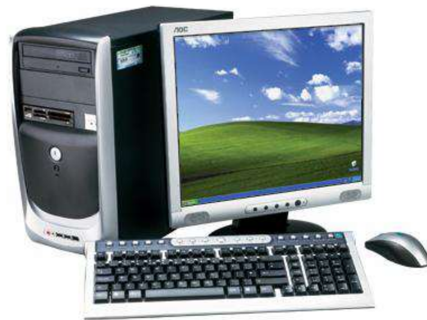


Fig. 2.1 *A Computer System*

(i) Motherboard

The main PCB (Printed Circuit Board) is sometimes alternatively known as a logical board or a main board of a Personal Computer. In fact, any complex electronic system is known as a motherboard. It includes a flat fibreglass platform which hosts the CPU (Central Processing Unit), the main electronic components, device controller chips, main memory slots, slots for attaching the storage devices and other subsystems. Figure 2.2 displays a motherboard.

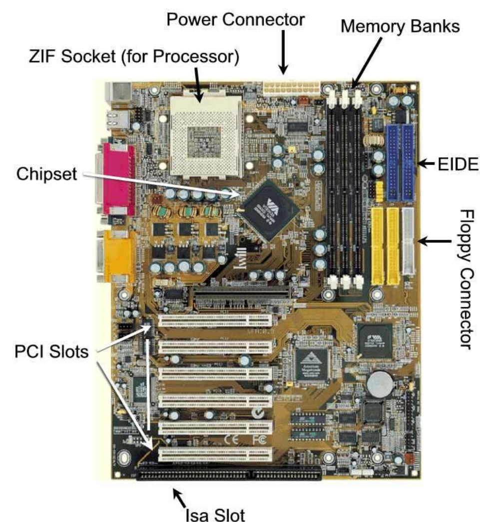


Fig. 2.2 *A Motherboard*

NOTES

(ii) Sockets and Ports

- **Main Power Socket:** The top part of the rear of the computer locates the main power cable socket, which supplies power from the electric mains to the computer system. This socket is the part of the main power supply unit of the computer.
- **Monitor Power Socket:** The socket that supplies the power from the computer system to the computer monitor is located below the main power cable socket. However, you might not find this socket in all computers and you can plug in the monitor directly in main power supply.
- **PS/2 Mouse Port:** Next you will find a small, round, green-colored port with seven holes and a small logo of the mouse printed next to it. This is where your PS/2 Mouse will be plugged in.
- **PS/2 Keyboard Port:** Right next to it you will find another similar purple-colored port with the keyboard logo printed next to it. This is where your PS/2 keyboard will be plugged in.
- **Fan Housings:** You will notice two fan housings at the back of your computer. One fan housing is a part of the power supply unit and the other will be somewhere below it to cool off the heat generated by the CPU.
- **Serial Ports:** It is a 9-pin connector normally used to attach the old serial port mouse, hand-held scanners, modems, joysticks, game pads, and other such devices.
- **Parallel Port:** It is a 25-pin connector used to attach parallel port printers, modems, external hard disk drives, etc.
- **Audio Jacks:** There are three audio jacks in your computer system. One jack is used for connecting your speakers or headphones, the second is used to connect the microphone and the third to connect to another audio device, such as a music system.
- **LAN Port:** The LAN port is where the RJ45 connector of your LAN cable is plugged in to connect your computer to other computers or the Internet.
- **USB Ports:** The USB port is designed to connect multiple peripheral devices in a single standardized interface and has a plug and play option that allows devices to be connected and disconnected without having to restart or turning off the computer. It has replaced many serial and parallel port devices such as mouse, printers, modems, joysticks, game pads, scanners, digital cameras, and other such devices.
- **VGA Port:** This is a 15-pin connector that connects the signal cable of the monitor to the computer.

Figure 2.3 displays monitor and CPU power cables and sockets.

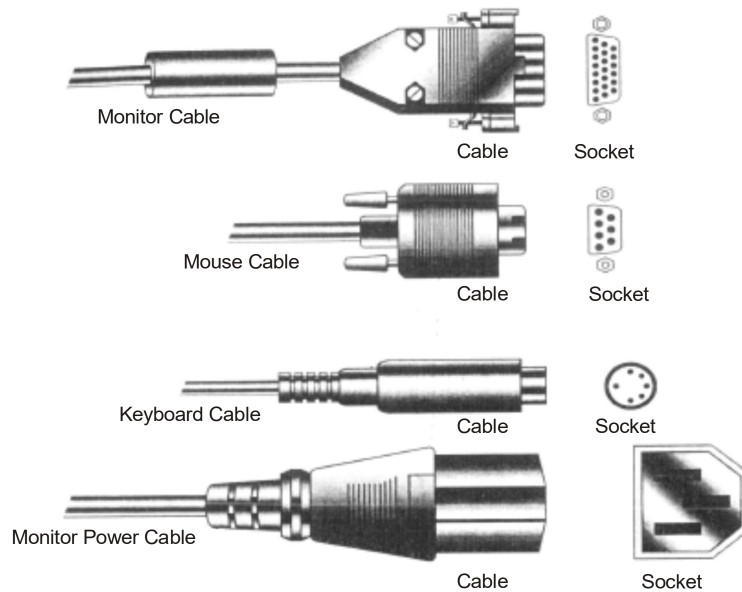


Fig. 2.3 *Monitor and CPU Power Cable and Sockets*

Figure 2.4 displays a LAN cable and a printer cable and its socket.

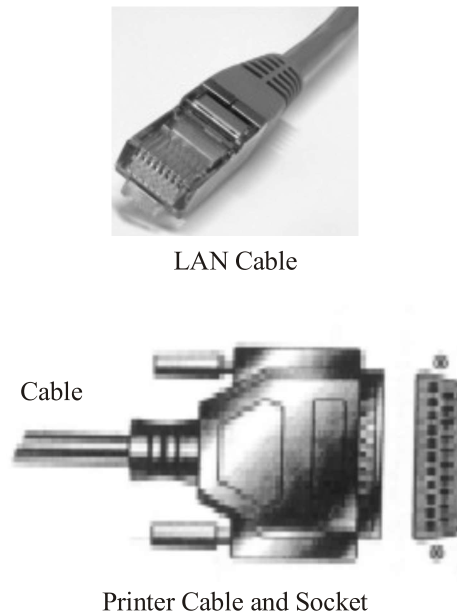


Fig. 2.4 *LAN Cable and Printer Cable with Socket*

2.2.1 CPU

The primary function of the computer is executing programs. The programs or the set of instructions are stored in the computer's main memory and are executed by the CPU. The CPU processes the set of instructions along with any calculations

and comparisons to complete the task. Additionally, the CPU controls and activates various other functions of the computer system. It also activates the peripherals to perform input and output functions. Figure 2.5 displays major components of the CPU.

NOTES

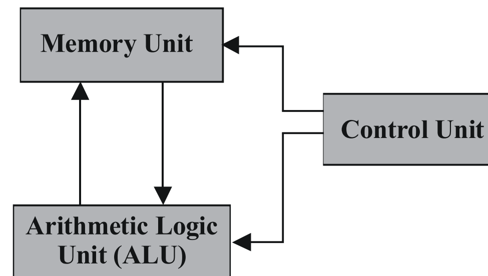


Fig. 2.5 *Major Components of a CPU*

Arithmetic and Logic Unit (ALU)

The Arithmetic and Logic Unit (ALU) of a computer system is the significant component where the actual execution of the instructions take place during the processing operations of instructions to provide desired output. All calculations are performed and all comparisons (decisions) are made in the ALU. ALU performs the actual computations as it consists of circuits that perform arithmetic operations, such as addition, subtraction, multiplication, division over data received from memory and it also used to compare numbers, such as less than, equal to, or greater than.

The data and instructions are stored in the primary storage before the processing is done by the ALU. At the time of processing, the data and instructions are transferred to the ALU. Remember that primary storage unit does not perform any processing. Therefore the data moves from primary storage to ALU and back again as storage several times before the processing is completed. After the completion of processing, the final results that are stored in the storage unit are sent to an output device.

While performing these operations the ALU takes data from the temporary storage are inside the CPU named as 'Registers'. Registers are a group of cells used for memory addressing, data manipulation and processing. All activities in the computer system are composed of thousands of individual steps which follow the predefined order in fixed intervals of time. These intervals are generated by the Clock Unit. Thus, every operation within the CPU takes place at the clock pulse. The clock rate is measured in MegaHertz (MHz) or GigaHertz (GHz).

Control Unit (CU)

The Control Unit (CU) of the computer system controls the functioning of the program instructions and various input and output devices. The control unit is responsible for maintaining the order of the instruction operations and then directing

the operation of the entire system by selecting, interpreting, and analysing the execution of the program instructions. Although, it does not perform any actual processing on the data, but the control unit acts as a central nervous system for the other components of the computer as it manages and coordinates the entire computer system. It obtains instructions from the program stored in main memory, interprets the instructions, and issues signals that cause other units of the system to execute them. Fundamentally, the control unit directs and controls the activities of the internal and external devices.

2.2.2 Memory

Storage and retrieval of instructions and data in a computer system is the responsibility of the ***memory***. In order to store data and instructions, the CPU constitutes many registers, though these are capable of storing very few bytes. All computers need storage space for temporarily storing instruction and data during the execution of the program as the CPU can process data at a speed that is much faster than the speed at which data can be transferred from disks to registers. This could lead to the CPU remaining free most of the time if the data was located in secondary storage including magnetic tapes and disks. The primary or the main memory is the temporary storage located in the computer hardware. Secondary storage or auxiliary memory constitutes devices that can give backup storage, such as magnetic tapes and disks. The memory is classified as follows:

(i) **Internal Processor Memory:** A small set of high-speed registers placed inside a processor and used for storing temporary data while processing.

(ii) **Primary Storage Memory:** The main memory of the computer which communicates directly with the processor. This memory is large in size and fast, but not as fast as the internal memory of the processor. It comprises a couple of integrated chips mounted on a printed circuit board plugged directly on the motherboard. Random Access Memory (RAM) is an example of primary storage memory.

(iii) **Secondary Storage Memory:** This stores all the system software and application programs and is basically used for data backups. It is much larger in size and slower than primary storage memory. Hard disk drives, floppy disk drives and flash drives are a few examples of secondary storage memory.

(iv) **Memory Capacity:** Capacity, in computers, refers to the number of bytes that it can store in its main memory. This is usually stated in terms of KiloBytes (KB) which is 1024 bytes or MegaBytes (MB) which is equal to 1024 KB (10,48,576 bytes). The rapidly increasing memory capacity of computer systems has resulted in defining the capacity in terms of GigaBytes (GB) which is 1024 MB (1,07,37,41,824 bytes). Thus a computer system having a memory of 256 MB is capable of storing (256 × 1024 × 1024) 26,84,35,456 bytes or characters.

2.2.3 Processors Used in PCs

The most significant part of the computer is the central processing unit or the CPU. The CPU is mostly a microprocessor-based chip located on a single or sometimes a multiple printed circuit boards and is an internal component of the system. It is directly connected to the motherboard; however, the compatibility of the mother board and the CPU depends on the specific series of the latter. Due to the tremendous amount of heat generated by the CPU, it contains a heat sink and a cooling fan.

Popular microprocessors include Intel and AMD, which manufacture IBM-compatible CPUs.

The brands of CPUs listed are not the only differentiating factors, between different processors. There are various technical aspects to these processors which allow us to differentiate between CPUs of different power, speed and processing capability. Accordingly, each of these manufacturers sells numerous product lines offering CPUs of different architecture, speed, price range, etc. The following are the most common aspects of modern CPUs that enable us to judge their quality or performance:

- **32 or 64-bit Architecture:** A bit is the smallest unit of data that a computer processes. 32 or 64-bit architecture refers to the number of bits that the CPU can process at a time.
- **Clock Rate:** The speed at which the CPU performs basic operations, measured in Hertz (Hz) or in modern computers MegaHertz – MHz or GigaHertz – GHz.
- **Number of Cores:** CPUs with more than one core are essentially multiple CPUs running in parallel to enable more than one operation to be performed simultaneously. Current ranges of CPUs offer up to eight cores. Currently, the Dual core (i.e., two cores) CPU is most commonly used for standard desktops and laptops and Quad core (i.e., four cores) is popular for entry level servers.
- **Additional Technology or Instruction Sets:** These refer to unique features that a particular CPU or range of CPUs offer to provide additional processing power or reduced running temperature. These range from Intel’s MMX, SSE3 and HT to AMD’s 3DNOW and Cool n Quiet.

These technical factors are the basic way to judge how a CPU will perform. It is important to consider multiple factors when looking at a CPU rather than just the clock speed or any one specification on its own. It is easy for a single-core processor to run music videos, Internet applications or games individually, but when multiple applications are run together, it starts to slow down. A system running on a dual-core processor would be able to multitask better then a single-core processor, while it is very easy for an 8-core processor to run all these applications plus a lot more without showing any signs of slowing down. However, Intel's 4-

core processors are actually two dual-core processors combined in a single processor, whereas AMD's 4-core processors are actually four processors built in a single chip.

A combination of the above-mentioned specifications, along with the operating systems that the processor supports and the specific purpose for which the computer is to be used, are the factors to be considered when deciding which CPU is the most suitable for your needs.

Check Your Progress

1. Which component of the computer is responsible for controlling different computer operations?
2. Name at least three ports of the motherboard along with their functions.

2.3 INPUT DEVICES

It is significant for computers to connect with the external world, i.e., receive and send data and information.

Computers have an input-output subsystem, referred to as I/O subsystem, which provides an efficient mode of communication between the central system and the outside world. Programs and data must be entered into the computer memory for processing, and results obtained from computations must be displayed or recorded for the user’s benefit. This data must first be stored in the computer’s memory after converting it into machine-readable form. The data will then be processed (average marks calculated) and sent from the memory to the output unit, which will present the data in a form that can be read by users.

The I/O devices that provide a means of communication between the computer and the outside world are known as peripheral devices. This is because they surround the CPU and the memory of a computer system. While input devices are used to enter data from the outside world into the primary storage, output devices are used to provide the processed results from primary storage to users.

Input devices are used to transfer user data and instructions to the computer. The most commonly used input devices can be classified into the following categories:

- Keyboard Devices (General and Special Purpose, Key-to-Tape, Key-to-Disk, Key-to-Diskette)
- Point-and-Draw Devices (Mouse, Trackball, Joystick, Light Pen, Touch Screen)
- Scanning Devices (Optical Mark Recognition, Magnetic Ink Character Recognition, Optical Barcode Reader, Digitizer, Electronic-Card Reader)
- Voice Recognition Devices
- Vision-Input Devices (Webcam, Video Camera)

NOTES

1. Keyboard Devices

Keyboard devices allow input into the computer system by pressing a set of keys mounted on a board, connected to the computer system. Keyboard devices are typically classified as general-purpose keyboards and special-purpose keyboards.

(i) General-Purpose Keyboard

The most popular keyboard used today is the 101-key with a traditional QWERTY layout, with an alphanumeric keypad, 12 function keys, a variety of special function keys, numeric keypad, and dedicated cursor control keys. It is so called because the arrangement of its alphanumeric keys in the upper-left row (as seen in the Figure 2.6).

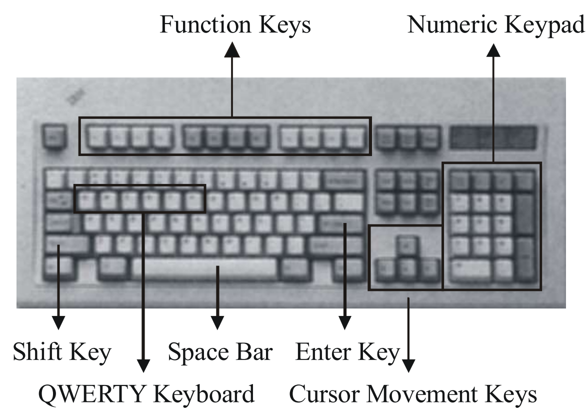


Fig. 2.6 *QWERTY Keyboard Layout*

- **Alphanumeric Keypad:** This contains keys for the English alphabets, 0 to 9 numbers, special characters like * + – / [ ], etc.
- **12 Function Keys:** These are keys labelled F1, F2 ... F12 and are a set of user- programmable function keys. The actual function assigned to a function key differs from one software package to another. These keys are also called soft keys since their functionality can be defined by the software.
- **Special Function Keys:** Special functions are assigned to each of these keys. The enter key, for example, is used to send the keyed-in data into the memory.
- **Numeric Keypad:** This consists of keys with numbers (0 to 9) and mathematical operators (+ – * /) defined on them. It is usually located on the right side of the keyboard and supports quick entry of numerical data.
- **Cursor Control Keys:** They are defined by the arrow keys used to move the cursor in the direction indicated by the arrow (top, down, left, right).

Another popular key arrangement, called Dvorak system, was designed for easy learning and use. It was designed with the most common consonants in one part and all the vowels on the other part of the middle row of the keyboard. This key arrangement made the users use alternate keystrokes back and forth between both the hands. This keyboard was never commonly used.

NOTES

(ii) Special-Purpose Keyboard

These have special-purpose keyboards to enable faster data entry and can be seen at the Automatic Teller Machines or the ATMs where the keyboard is required for limited functionality (support for some financial transactions) by the customers. These keyboards are specifically designed for special types of applications only.

(iii) Key-to-Tape, Key-to-Disk, Key-to-Diskette

Used for data entry only; these processor-based workstations normally have a keyboard and a small monitor. The function of the processor is to check the accuracy of the data when it is being entered.

The screen displays data as it is being entered. These facilities are very useful and desirable during mass data entry and are therefore becoming very popular in data processing centers.

2. Point-and-Draw Devices

The keyboard facilitates input of data only in text form. While working with display-based packages, we usually point to a display area and select an option from the screen (fundamentals of GUI applications). For such cases, the sheer user-friendliness of input devices that can rapidly point to a particular option displayed on screen and support its selection resulted in the advent of various point-and-draw devices.

(i) Mouse

A mouse is a small input device used to move the cursor on a computer screen to give instructions to the computer and to run programs and applications. It can be used to select menu commands, move icons, size windows, start programs, close windows, etc. Nowadays, the mouse is the most important device in the functioning of a Graphical User Interface (GUI) of almost all computer systems. Figure 2.7 displays an optical mouse.



Fig. 2.7 *Optical Mouse*

You can click a mouse button, i.e., press and release the left mouse button, to select an item. You can *right click*, i.e., press and release the right mouse button to display a list of commands. You can *double click*, i.e., quickly press the left mouse button twice without any time gap between the press of the buttons, to open a program or a document. You can also *drag and drop*, i.e., place the cursor over an item on the screen and then press and hold down the left mouse button. Holding down the button, move the cursor to where you want to place the item and then release the button.

NOTES

(ii) Touch Screen

A touch screen is probably one of the simplest and most intuitive of all input devices. It uses optical sensors in or near the computer screen that can detect the touch of a finger on the screen. Once the user touches a particular screen position, sensors communicate the position to the computer. This is then interpreted by the computer to understand the user's choice for input. The most common usage of touch screens is in information kiosks where users can receive information at the touch of a screen. Figure 2.8 displays a touch screen.



Fig. 2.8 *A Touch Screen*

(iii) Touch Pads

Input devices that are touch sensitive and take user input to manage onscreen pointer and other functions are known as touch pads. The touch pad enables the user to interact with the screen through the use of fingers that are dragged across the pad in various positions. The touch pad is used in place of the mouse which is an external peripheral device. Touch pads are usually present in notebooks and laptops as they provide convenience and space. Touch pads are touch as well as pressure sensitive and respond to finger drag and tapping combinations. In addition to the normal functions of a mouse, a user can also perform the scroll function by moving the finger at certain points, (the top and bottom corners of the right side of the pad). Touch pads are more convenient than a mouse as they put less pressure on the wrist and hand. The functionality of a touch pad is not limited by the manufacturer's user interface. They can be user programmed to recognize a combination of finger and tap moments to perform new actions as an input device. Touch pads have shown steady growth in market demand and user acceptability over all phases of their development cycles. Their growth is further expected to continue in the same way as the demand for touch-sensitive portable devices with more functionality and better appearance is on the rise. Figure 2.9 displays a touch pad.

NOTES

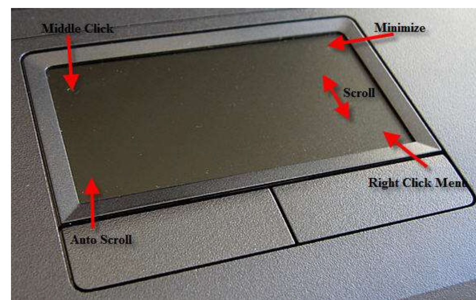


Fig. 2.9 A Touch Pad

(iv) Light pen

A light pen is a small input device used to select and display objects on a screen. It functions with a light sensor and has a lens on the tip of a pen-shaped device. The light receptor is activated by pointing the light pen towards the display screen and it then locates the position of the pen with the help of a scanning beam application to directly draw on screen. Figure 2.10 displays a light pen.

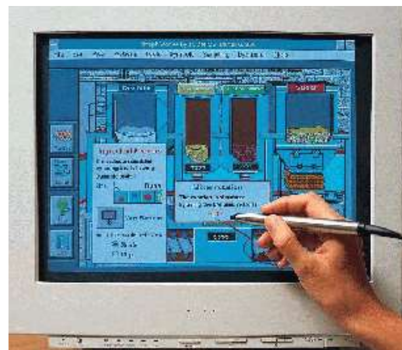


Fig. 2.10 A Light Pen being used to Display Objects on Screen

(v) Trackball

A trackball is a pointing device that is much like an inverted mouse. It consists of a ball inset in a small external box or adjacent to, and in the same unit, as the keyboard of some portable computers. Figure 2.11 displays a trackball.



Fig. 2.11 A Trackball

It is more convenient and requires much less space than the mouse since here the whole device is not moved (as in the case of a mouse). Trackball comes in various

shapes but supports the same functionality. Typical shapes used are a ball, a square and a button (typically seen in laptops).

NOTES

(vi) Joystick

A joystick is a vertical stick that moves the graphic cursor in the direction the stick is moved. It consists of a spherical ball, which moves within a socket and has a stick mounted on it. The user moves the ball with the help of the stick that can be moved left or right, forward or backward, to move and position the cursor in the desired location. Joysticks typically have a button on top that is used to select the option pointed by the cursor. Figure 2.12 displays a joystick.



Fig. 2.12 *A Joystick*

Video games, training simulators and control panels of robots are some common activities where joysticks are used.

3. Scanning Devices

Scanning devices are input devices used for direct data entry from the source document into the computer system. With the help of the scanner you can capture your images and documents and convert it into digital formats for easy storage on your computer. The scanner comprises of two major components, the first component lights up the page in order to capture the optical image and the second component converts that captured image into a digital format for viewing and storing it on your computer system. Figure 2.13 displays a hand-held scanner.



Fig. 2.13 *A Hand-Held Scanner*

There are two types of scanners, contact and laser. Both illuminate the image first to calculate the reflected light and determine the value of the captured image. Hand-held contact scanners make contact as they are brushed over the printed matter to be read. Laser-based scanners are more versatile and can read data passed near the scanning area. Figure 2.14 displays a flat-bed scanner.



Fig. 2.14 *A Flat-Bed Scanner*

Hand-held scanners are used where the information to be scanned or the volume of documents to be scanned is very low. They are much cheaper as compared to the flat-bed scanners. Capturing information using scanners reduces the possibility of human error typically seen during large data entry. The reduction in human intervention improves the accuracy of data and provides for timeliness of the information processed.

Source data automation is the recent development for data input technologies. Source data automation does not require any manual data entry; rather, it captures the data as a derivative of the routine business activity.

(i) Optical Mark Recognition (OMR)

The OMR devices can scan marks from a computer-readable paper. Such devices are used by universities and institutes to mark test sheets where the candidate selects and marks the correct answer from multiple choices given on a special sheet of paper. These marksheets are not required to be evaluated manually as they are fed in the OMR and the data is then transferred to the computer system for further evaluation. Figure 2.15 displays an OMR sheet.

Answer Sheet

1. a b c d
2. a b c d
3. a b c d
4. a b c d

Page 1 of 4

Fig. 2.15 *Sheet Read by an OMR Device*

NOTES

The actual technique used by an OMR device once again involves focussing a light on the page being scanned, thereby detecting the reflected light pattern for the marks. Pencil marks made by the user reflect the light determining which responses are marked.

(ii) Magnetic Ink Character Recognition (MICR)

Magnetic ink character recognition is like an optical mark recognition device and is used only in the banking industry. MICR devices scan cheque numbers directly from the cheque leaflets and then automatically feed them in the computer systems for further use, doing the job quickly, accurately and efficiently. Figure 2.16 displays magnetic ink characters on a bank cheque.



Fig. 2.16 *A Bank Cheque using MICR Technology*

Banks using MICR technology print cheque books on special types of paper. The necessary details of the bank (like the bank's identification code, relevant account number and cheque number) are pre-printed on the cheques using an ink that contains iron oxide particles that can be magnetized.

MICR readers are used to read and sort cheques and deposits. An MICR reader-sorter reads the data on the cheques and sorts the cheques for distribution to other banks and customers or for further processing.

(iii) Optical Bar Code Reader (OBR)

Data coded in the form of small vertical lines forms the basis of bar coding. Alphanumeric data is represented using adjacent vertical lines called *bar codes*. These are of varying widths and the spacing between them are used to uniquely identify books, merchandise in stores, postal packages, etc. Figure 2.17 shows a bar code used on a book for its unique identification.

NOTES



Fig. 2.17(a) *An Example of a Bar Code*



Fig. 2.17(b) *A Bar Code Reader*

A bar code reader uses laser beam technology. The laser beam is moved across the pattern of bars in a bar code. These bars reflect the beam in different ways. The reflected beam is then sensed by a light-sensitive detector, which then converts the light patterns into electrical pulses, thereby transmitting them to logic circuits for further conversion to alphanumeric value. Bar code devices are available as hand-held devices.

(iv) Digitizer

Digitizers are used to convert drawings or pictures and maps into a digital format for storage into the computer. A digitizer consists of a digitizing or graphics tablet, which is a pressure-sensitive tablet, and a pen with the same X and Y coordinates as on the screen. Some digitizing tablets also use a crosshair device instead of a pen. The movement of the pen or crosshair is reproduced simultaneously on the display screen. When the pen is moved on the tablet, the cursor on the computer screen moves simultaneously to the corresponding position on the screen (X and Y coordinates). This allows the user to draw sketches directly or input existing sketched drawings easily. Digitizers are commonly used by architects and engineers as a tool for Computer-Aided Designing (CAD). Figure 2.18 displays a digitizing tablet.



Fig. 2.18 *A Digitizing Tablet*

(v) Electronic-Card Reader

Card readers are devices that also allow direct data input into a computer system. The electronic-card reader is connected to a computer system and reads the data

encoded on an electronic card and transfers it to the computer system for further processing. Electronic cards are plastic cards with data encoded on them and meant for a specific application. A popular example of electronic cards are the plastic cards issued by banks to their customers for use in ATMs. Electronic cards are also used by many organizations for controlling access of various types of employees to physically secured areas. Figure 2.19 displays an access card security system.



Fig. 2.19 *An Access Card Security System*

Based how the data is encoded, electronic cards may be either magnetic strip cards or smart cards. Magnetic strip cards have a magnetic strip on the back of the card. Data stored on magnetic strips cannot be read with the naked eye, a useful way to maintain confidential data. Smart cards, going a stage further, have a built-in microprocessor chip where data can be permanently stored. They also possess some processing capability making them suitable for a variety of applications. To gain access, for example, an employee inserts a card or badge in the reader. This device reads and checks the authorization code before permitting the individual to enter a secured area. Since smart cards can hold more information as compared to magnetic strip cards, they are gaining popularity.

4. Voice Recognition Devices

Voice recognition devices consist of a microphone attached to the computer system. A user speaks into the microphone to input data. The spoken words are then converted into electrical signals (this is in the analog form). A digital-to-analog converter then converts the analog form to digital form (0s and 1s) that can be interpreted by the computer. The digitized version is then matched with the existing pre-created dictionary to perform the necessary action. Voice recognition devices have limited usage today because they have several problems. Not only do they require the ability to recognize who is speaking, but also what is being said (the

NOTES

message). This difficulty arises primarily because people speak with different accents and different tones and pitches. The computer requires a large vocabulary to be able to interpret what is being said. Thus, voice recognition systems are successful in a limited domain.

5. Vision-Input Devices

Vision input devices allow data input in the form of images. It mostly consists of a digital camera, which focuses on the object whose picture is to be taken. The camera creates the image of the object in digital format, which can then be stored within the computer.

The speech recognition system digitizes the voice input and this system similarly compares the digitized images to be interpreted to the prerecorded digitized images of the database of your computer system. After finding the right match, it sends it for further processing or pre-defined action.

(i) Web Camera

A web camera is a video capturing device attached to the computer system, mostly using a USB port used for video conferencing, video security, as a control input device. Figure 2.20 displays a web camera.



Fig. 2.20 *A Web Camera*

Check Your Progress

3. What are the categories in which the input devices are classified?
4. What are scanning devices?

2.4 OUTPUT DEVICES

An electromechanical device that accepts data from the computer and translates it into a form that can be understood by the external world is known as an output device. The processed data, stored in the memory of the computer, is sent to an output unit, which then transforms the internal representation of data into a form that can be read by the users.

Normally, the output is produced on a display unit like a computer monitor or can be printed through a printer on paper. At times, speech outputs and mechanical outputs are also used for some specific applications.

Output produced on display units or speech output that cannot be touched is referred to as softcopy output while output produced on paper or material that can be touched is known as hardcopy output. A wide range of output devices are available today and can be broadly classified as follows:

- Display Devices (Monitors, Multimedia Projectors)
- Speakers
- Printers (Dot Matrix, Inkjet, Laser)
- Plotters (Flatbed, Drum)

1. Display Devices

A display device is the most essential peripheral of a computer system. Initially, alphanumeric display terminals were used that formed a 7×5 or 9×7 array of dots to display text characters only. As a result of the increasing demand and use of graphics and GUIs, graphic display units were introduced. These graphic display units are based on series of dots known as pixels used to display images. Every single dot displayed on the screen can be addressed uniquely and directly.

Since each dot can be addressed as a separate unit, it provides greater flexibility for drawing pictures. Display screen technology may be of the following categories:

(i) **Cathode Ray Tube (CRT):** The CRT consist of an electron gun with an electron beam controlled with electromagnetic fields and a phosphate-coated glass display screen structured into a grid of small dots known as pixels. The image is created with the electron beam produced by the electron gun, which is thrown on the phosphor coat displayed by the electromagnetic field.

(ii) **Liquid Crystal Display (LCD):** Liquid Crystal Display (LCD) was first introduced in the 1970s in digital clocks and watches, and is now widely being used in computer display units. The Cathode Ray Tube (CRT) was replaced with the Liquid Crystal Display (LCD) making it slimmer and more compact. However, the image quality and the image color capability got comparatively poorer.

The main advantage of LCD is its low energy consumption. It finds its most common usage in portable devices where size and energy consumption are of main importance.

(i) **Projection Display:** Projection display technology is characterized by replacing the personal size screen with large screens upon which the images are projected. It is attached to the computer system and the magnified display of the computer system is projected on a large screen.

NOTES

(ii) **Monitors:** Monitors use a Cathode Ray Tube (CRT) to display information. It resembles a television screen and is similar to it in other respects as displayed in Figure 2.21.

The monitor is connected to a keyboard for manual input of characters. The screen displays information as it is keyed in, enabling a visual check of the input before it is transferred to the computer. It is also used to display the output from the computer and hence serves as both an input and an output device.



Fig. 2.21 *A CRT Monitor*

The monitor is the most commonly used input/output device today and is also known as a soft copy terminal. A printing device is usually required to provide a hard copy of the output.

2. Multimedia Projectors

A multimedia projector is an output device which is used to project information from the computer onto a large screen so that it can be viewed by a large group of people. Prior to this, the standard mode of making presentations was to make transparencies and project them using an overhead projector. This was a tedious and time-consuming activity, since for every change in the subject matter a new transparency had to be prepared. And of course, since electronic cut, copy and paste was not possible, this meant additional work.

A multimedia projector can directly be plugged into a computer system and the information projected on a large screen thereby making it possible to present information to a large audience. The presenter can also use a pointer to emphasize specific areas of interest in the presentation.

LCD flat screens connected to the computer systems for projecting the LCD image using an overhead multimedia projector are widely used today.

Speakers

Speakers are used to produce music or speech from programs, a speaker port (a port is a connector in your computer wherein you can connect an external device) allows connection of a speaker to a computer. Speakers can be built into the computer or can be attached separately. Figure 2.22 displays speakers.



Fig. 2.22 Speakers

3. Printers

Printers are used for generating output printed on paper. There is a huge range of commercially available printers today (estimated to be 1500 different types). These printers can be classified into categories based on the following:

- Printing Technology
- Printing Speed
- Printing Quality

(i) **Printing Technology:** Printers can be classified as impact or non-impact printers, based on the technology they use for producing output. Impact printers work on the mechanism similar to a manual typewriter where the printer head strikes on the paper and leaves the impression through an inked ribbon. Dot matrix printers and character printers fall under this category. Non-impact printers use chemicals, inks, toners, heat or electric signals to print on the paper and they do not physically touch the paper while printing.

(ii) **Printing Speed:** The number of characters printed in a unit of time reflects the printing speed. Based on speed, these may be classified as character printer (prints one character at a time), line printers (prints one line at a time), and page printers (print the entire page at a time). Printer speed is measured in terms of characters-per-second or *cps* for a character printer, lines-per-minute or *lpm* for a line printer, and pages-per-minute or ppm for a page printer.

(iii) **Printing Quality:** The quality determined by the resolution of printing and is characterized by the number of dots that can be printed per linear inch, horizontally or vertically. It is measured in terms of dots-per-inch or *dpi*. Printers can be classified as near-letter-quality or NLQ, letter-quality or LQ, near-typeset-quality or NTQ and typeset-quality or TQ, based on their printing quality. NLQ printers have resolutions of about 300 dpi, LQ of about 600 dpi, NTQ of about 1200 dpi and TQ of about 2000 dpi. NLQ and LQ printers are used for ordinary printing in day-to-day activities, while NTQ and TQ printers are used to produce top-quality printing, typically required in the publishing industry.

The different types of printers are as follows:

(i) Dot Matrix

Dot matrix printers are the most widely used impact printers in personal computing. These printers use a print head consisting of a series of small metal pins that strike on a paper through an inked ribbon, leaving an impression on the paper through the ink transferred. Characters thus produced are in a matrix format. The shape of each character, i.e., the dot pattern, is obtained from information held electronically. Figure 2.23 displays characters formed using dots and dot matrix printer.

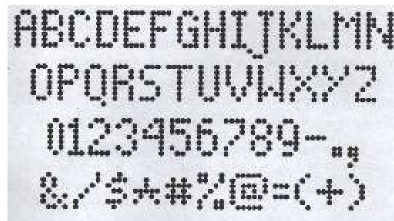


Fig. 2.23(a) *Characters Formed using Dots*

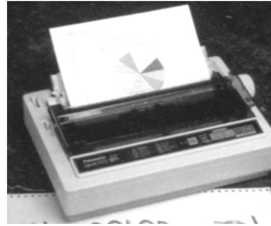


Fig. 2.23(b) *Dot Matrix Printer*

The speed, versatility and ruggedness, combined with low cost, tend to make such printers particularly attractive in the personal computer market. Typical printing speeds in case of dot matrix printers range between 40–1000 characters per second (cps). In spite of all these features in dot matrix printer technology, the low print quality gives it a major disadvantage.

(ii) Inkjet

Inkjet printers are based on the use of a series of nozzles for propelling droplets of printing ink directly on almost any size of paper. They, therefore, fall under the category of non-impact printers. The print head of an inkjet printer consists of a number of tiny nozzles that can be selectively heated up in a few microseconds by an IC register. When this happens, the ink near it vaporizes and is ejected through the nozzle to make a dot on the paper placed in front of the print head. The character is printed by selectively heating the appropriate set of nozzles as the print head moves horizontally. Figure 2.24 displays an inkjet printer.



Fig. 2.24 *An Inkjet Printer*

NOTES

(iii) Laser

Laser printers work on the same printing technology as photocopiers, using static electricity and heat to print with a high quality powder substance known as toner. Figure 2.25 displays a laser printer.



Fig. 2.25 *A Laser Printer*

Laser printers are capable of converting computer output into print, page by page. Since characters are formed by very tiny ink particles, they can produce very high quality images (text and graphics). They generally offer a wide variety of character fonts, and are silent and fast in use. Laser printers are faster in printing speed than other printers discussed above. Laser printers can print from 10 pages to 100 pages per minute, depending upon the make/model. Laser is high quality, high speed, high volume and non-impact technology that works on almost any kind of paper. Even though this technology is more expensive than inkjet printers, it is preferred because of its unmatched features, such as high quality, high-speed printing and noiseless and easy-to-use operations.

4. Plotters

Plotters are used to make line illustrations on paper. They are capable of producing charts, drawings, graphics, maps and so on. A plotter is much like a printer but is designed to print graphs instead of alphanumeric characters. Based on the technology used, there are mainly two type of plotters: *pen plotters* and *electrostatic plotters*. Pen plotters, also known as flatbed plotters, draw images with multicolored pens attached to a mechanical arm. Electrostatic plotters, also known as drum plotters, work on the same technology as laser printers.

Flatbed plotters and drums are the most commonly used plotters.

(i) Flatbed Plotters

Flatbed plotters have a flat base like a drawing board on which the paper is laid [as shown in Figure 2.26(a)]. One or more arms, each of them carrying an ink pen, moves across the paper to draw. The arm movement is controlled by a microprocessor (chip). The arm can move in two directions, one parallel to the

NOTES

plotter and the other perpendicular to it (called the *x* and *y* directions). With this kind of movement, it can move very precisely to any point on the paper placed below.

The computer sends the commands to the plotter which are translated into *x* and *y* movements. The arm moves in very small steps to produce continuous and smooth graphics. The size of the plot in a flatbed plotter is limited only by the size of the plotter's bed. The advantage of flatbed plotters is that the user can easily control the graphics. He can manually pick up the arm anytime during the production of graphics and place it on any position on the paper to alter the position of graphics to his choice.

However, flatbed plotters occupy a large amount of space.

(ii) Drum Plotters

Drum plotters move the paper with the help of a drum revolver during printing [as shown in Figure 2.26(b)]. The arm carrying a pen moves only in one direction, perpendicular to the direction of the motion of the paper. It means that while printing, the plotter pens print on one axis of the paper and the cylindrical drum moves the paper on the other axis. With this printing technology, the plotter has an advantage to print on unlimited length of paper but on a limited width. Drum plotters are compact and lightweight compared to flatbed plotters. This is one of the advantages of such plotters.

The disadvantage of drum plotters, however, is that the user cannot freely control the graphics when they are being created. Plotters are more expensive when compared to printers. Typical application areas for plotters include Computer-Aided Engineering (CAE) applications, such as Computer-Aided Design (CAD) and Computer-Aided Manufacturing (CAM), architectural drawing and map drawing.

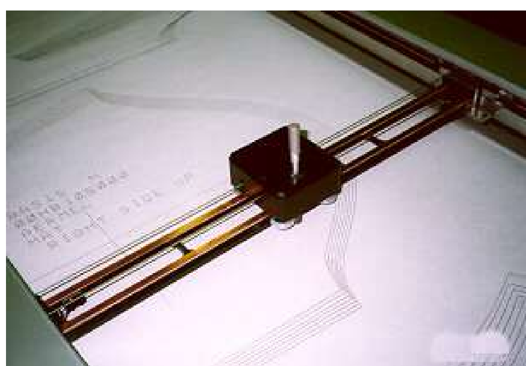


Fig. 2.26(a) *Top View of a Flatbed Plotter*



Fig. 2.26(b) *A Drum Plotter*

Check Your Progress

5. How is an output derived in computer?
6. How does a laser printer work?

2.5 STORAGE DEVICES

NOTES

The most common properties used for characterizing and evaluating the storage unit of the computer system are expressed below:

1. **Storage Capacity:** Represents the size of the memory. It is the amount of data that can be stored in the storage unit. Primary storage units have less storage capacity as compared to secondary storage units. The capacity of internal memory and main memory can be expressed in terms of the number of words or bytes.The capacity of external or secondary storage, on the other hand, is measured in terms of bytes.
2. **Storage Cost:** Cost is another key factor that is of prime concern in a memory system . It is usually expressed per bit. It is obvious that lower costs are desirable. It is worth noting that with the increase in the access time for memories , the cost decreases.
3. **Access Time:** The time required to locate and retrieve the data from the storage unit. It is dependant on the physical characteristics and access mode used for that device. Primary storage units have faster access time as compared to secondary storage units.
4. **Access Mode:** Memory comprises various locations. Access mode is the mode in which information is accessed from the memory. The user can access memory devices in any of the following ways:
   (a) **Random Access Memory (RAM):** This refers to the mode in which any memory location can be accessed in any order in the same amount of time. Ferrite and semiconductor memories, which usually constitute the primary storage or main memory, are of this nature.
   (b) **Sequential Access:** Memories that can be accessed only in a pre-defined sequence are sequential access memories. Here, since sequencing through other locations precedes the arrival at a desired location, the access time varies according to the location. Information on a sequential device can be retrieved in the same sequence in which it was stored. Songs stored on a cassette, that can be accessed only one by one, is an example of sequential access. Typically, magnetic tapes are sequential access memory.
   (c) **Direct Access:** Sometimes, the information is neither accessed randomly nor in sequence but something in between. In this type of access, a separate read/write head exists for each track, and on a track , you can access the information serially. This semi-random mode of access exists in magnetic disks.
5. **Permanence of Storage:** If the storage unit can retain the data even after the power is turned off or interrupted, it is termed as non-volatile storage. And, if the data is lost once the power is turned off or interrupted, it is called

volatile storage. It is obvious from these properties that the primary storage units of the computer systems are volatile, while the secondary storage units are non-volatile. A non-volatile storage is definitely more desirable and feasible for storage of large volumes of data.

2.5.1 Primary Storage Devices

1. Static and Dynamic RAM

The main memory is the central storage unit in a computer system. It is a relatively large and fast memory. It is used to store programs and data during computer operations. The principal technology used for the main memory is based on semiconductor-integrated circuits. There are two possible modes in which the integrated circuit RAM chips are available. These modes are: *Static* and *Dynamic*.

The Static RAM (SRAM) stores binary information using clocked sequential circuits. The stored information remains valid only as long as power is applied to the unit. On the other hand, Dynamic RAM (DRAM) stores binary information in the form of electric charges that are applied to capacitors inside the chip. The stored charge on the capacitors tends to discharge with time and so must be periodically recharged by refreshing the dynamic memory. The Dynamic RAM offers larger storage capacity and reduced power consumption. Therefore, large memories use dynamic RAM, while static RAM is mainly used for specialized applications.

The different types of memory discussed above are both of the read/write type. What about a memory where only one of the operations is possible, e.g., if we allow only reading from the memory (cannot change the information in the memory)? The memory might have some major importance; like an important bit of the computer's operating system which normally does not change can be stored in this kind of memory. Such a memory is called ROM (Read Only Memory).

2. Read-Only Memory (ROM)

Most of the memory in a general-purpose computer is made of RAM integrated circuit chips, but a portion of the memory may be constructed using ROM chips. Originally, RAM was used to refer to random-access memory, but now we use the term read/write memory to distinguish it from Read-Only Memory (since ROM is also random access). RAM is used for storing the bulk of the programs and data that are subject to change, while ROM is used to store programs that are permanently resident in the computer and do not change once the production of the computer is completed.

Among other things, the ROM portion of the main memory is used for storing an initial program called the bootstrap loader, whose function is to get the computer software operating when power is turned on. Since RAM is volatile, its contents are destroyed when power is turned off. The contents of ROM remain unchanged even after the power is turned off and on again.

NOTES

Read-only memories can be manufacturer-programmed or user-programmed. When the data is burnt into the circuitry of the computer by the manufacturer, it is called manufacturer-programmed ROM; for example, a personal computer manufacturer may store the boot program permanently in the ROM chip of the computers manufactured by it. Such chips are supplied by the manufacturer and are not modifiable by users. This is an inflexible process and requires mass production. Thus, a new kind of ROM, known as Programmable Read-Only Memory (PROM), was designed. This is also non-volatile in nature. It can be written only once using some special equipment. The supplier or the customer can electrically perform the writing process in PROM.

In both ROM and PROM, you can perform write operations only once and you cannot change whatever you have written. But what about the cases where you mostly read but also write a few times? Another type of memory chip called EPROM (Erasable Programmable Read-Only Memory) was developed to take care of such situations. EPROMs are typically used by R&D personnel who experiment by changing micro-programs on the computer system to test their efficiency.

Further, EPROM chips are of two types: EEPROMs (Electrically EPROM) in which high voltage electric pulses are used to erase stored information, and UVEPROM (Ultra Violet EPROM) in which stored information is erased by exposing the chip for a while to ultraviolet light.

Figure 2.27 summarizes the various types of Random Access Memories.

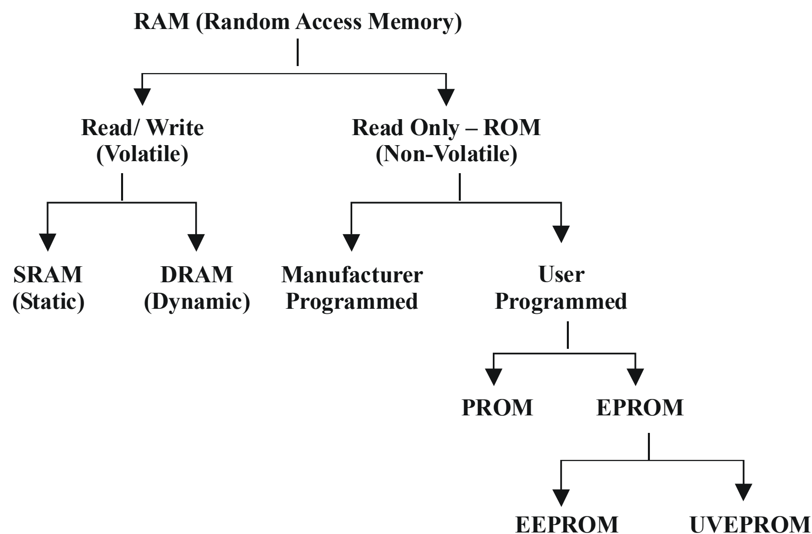


Fig. 2.27 *Types of Random Access Memories*

2.5.2 Secondary Storage Devices

As discussed earlier, RAM is a volatile memory with limited storage capacity. The cost of RAM is also relatively higher as compared to secondary memory. Logic dictated that a relatively cheaper media, showing some sort of permanence of

NOTES

storage, be used. As a result, additional memory called *external* or *auxiliary memory* or *secondary storage* is used in most computers.

The magnetic medium was found to be long lasting and fairly inexpensive, and therefore, became an ideal choice for large storage requirements. The use of magnetic tapes and disks as storage media is very common. As optical technology is advancing, optical disks are turning out to be one of the major secondary storage devices.

1. Cache Memory

Cache memories are small, fast memories placed between the CPU and the main memory. They are faster than the main memory with access times closer to the speed of the CPU. Caches are fast, but very expensive. So, they are used only in small quantities. As an example, caches of size 64K, 128K are normally used in PC-386 and PC-486, which can have 1 to 8 MB of RAM or even more. Cache memories provide fast speed memory retrieval without compromising the size of the memory.

If the memory is so small, would it be advantageous to increase the overall speed of memory? This can be answered with the help of a phenomenon known as locality of reference. Let us examine what this means.

If we analyse a large number of typical programs,we would find that memory references at any given interval of time tend to be confined to a few localized areas in the memory. This phenomenon is called the property of locality of reference. This is true because most of the programs typically contain iterative loops (like 'for' or 'while' loops). During the execution of such programs, the same set of instructions (within the loop) are executed many times. The CPU repeatedly refers to the set of instructions in the memory that constitute the loop. Every time a specific subroutine is called, its set of instructions is fetched from the memory. Thus loops and subroutines tend to localize the references to memory for fetching instructions.

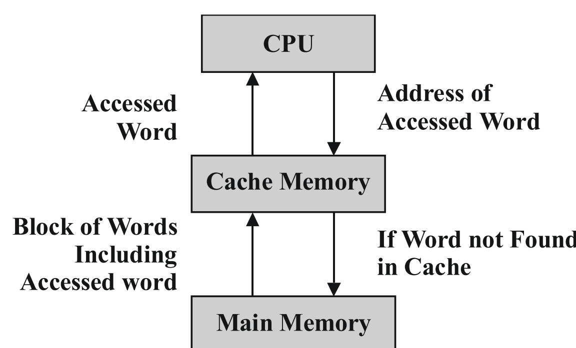


Fig. 2.28 *Functioning of the Cache Memory*

Based on the locality of reference, we understand that the cache has a copy of certain portions of main memory. First, the memory read or write operation is checked with the cache, and in case of availability of desired data in the cache, it

is used directly by the CPU. Otherwise, a block of words is read from main memory to cache and the word is then used by the CPU from cache. Figure 2.28 explains the functioning of the cache memory.

NOTES

3. Magnetic Tapes

Magnetic tapes are used for storing files of data that are sequentially accessed or not used very often and are stored off-line. They are typically used as backup storage for archiving of data.

In case of magnetic tapes, a tape (plastic ribbon usually 1/2 inch or 1/4 inch wide and 50 to 2400 feet long) is wound on a spool and its other end is threaded manually on a take-up spool. The Beginning Of the Tape (BOT) is indicated by a metal foil called a *marker*. When a write command is given, a block of data (records are usually grouped in blocks of two or more) is written on the tape. The next block is then written after a gap (called Inter Block Gap or IBG). A series of blocks are written in this manner. The End Of Tape (EOT) is indicated by an end-of-tape marker which is a metal foil stuck in the tape. After the data is written, the tape is rewound and kept ready for reading. Figure 2.29(a) shows the data organization on a magnetic tape. Figure 2.29(b) and 2.29(c) show magnetic tape reels and magnetic tape cartridges.

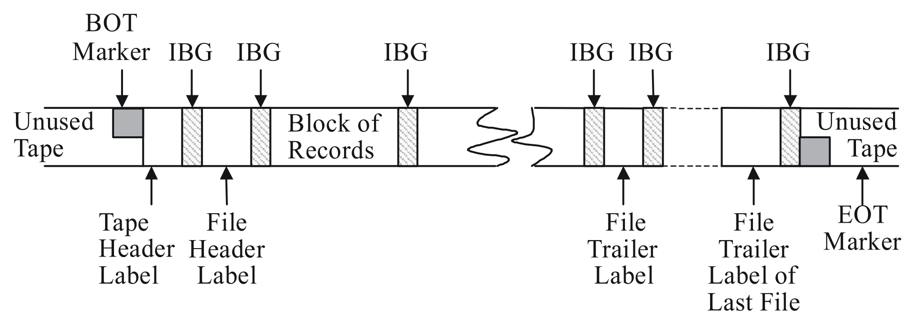


Fig. 2.29(a) *Data Organization on a Magnetic Tape*



Fig. 2.29(b) *Magnetic Tape Reel*



Fig. 2.29(c) *Magnetic Tape Cartridge*

The tape is read sequentially, i.e., data can be read in the order in which the data has been written. This implies that if the desired record is at the end of the tape, all the earlier records have to be read before it is reached. A typical example of a tape can be seen in a music cassette, where, to listen to the fifth song one must listen to, or traverse, the earlier four songs. The access time of information stored on tape is therefore very high as compared to that stored on a disk.

The storage capacity of the tape depends on its data recording density and the length of the tape. The data recording density is the amount of data that can be stored or the number of bytes that can be stored per linear inch of tape. The data recording density is measured in BPI (Bytes Per Inch).

Thus,

Storage capacity of a tape = Data recording density × Length of tape

It is worth noting that the actual storage capacity for storing user data is much less owing to the file header labels, file trailer labels, BOT and EOT markers, and the use of IBGs.

Some commonly used magnetic tapes are:

- 1/2 inch Tape Reel
- 1/2 inch Tape Cartridge
- 1/4 inch Streamer Tape
- 4 mm DAT (Digital Audio Tape) – Typical Capacity of 4GB to 14 GB

4. Magnetic Disks

Magnetic disks are direct-access medium, and so are the most popular online secondary storage devices. Direct-access devices are also called random-access devices because information is literally available at random or in any order. There is direct access to any location on the device. Thus, approximately equal access time is required for each location. An example of this is a music CD where if you wish to listen to the fifth song, you can directly select the fifth track without having to fast forward the previous four.

NOTES

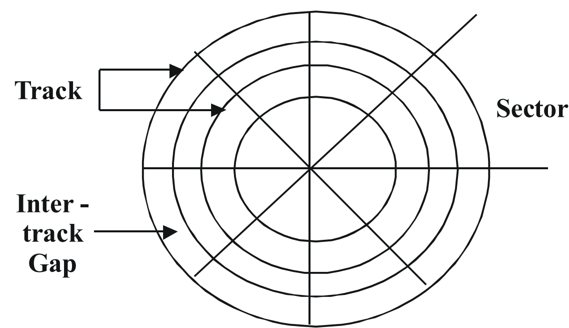


Fig. 2.30 Logical Layout of a Magnetic Disk

A magnetic disk refers to a circular plate made of metal or plastic and coated with magnetized material (ss shown in Figure 2.30). Often both sides of the disk are used. Data is recorded on the disk in the form of magnetized and non-magnetized spots (not visible to the naked eye) representing 1s and 0s.

(i) Disk Devices

A disk drive is a peripheral device used to store and collect information. It can be removable or fixed, high capacity or low capacity, fast or slow speed, and magnetic or optical.

Structurally, a drive is the object inside which a disk(s) is either permanently or temporarily stored. While a disk contains the media on which the data is stored, a drive contains the machinery and circuitry required for implementing the read / write operations on the disk.

The disk looks literally like a flat circular disk. The computer writes information onto the disk, where it is stored in the same form as it is stored on a cassette tape. Disks, as such, are just magnetically coated rolls or circular disks which are divided into sectors and tracks. The data is accordingly stored and numbered with respect to the track and sector number on the disk. Only the structure of the media is different from one to another. Examples of removable disk drives are DVD, CD-ROM, floppy disk drive, etc. The examples of non-removable disk drives include hard disk.

The method of accessing data could be sequential access (Magnetic Tape Drives) or random access (HDD, DVD), where the read/write head can directly go to any location on the disk and perform action.

(ii) Diskettes

Based on the size and packaging of the disks, they can be classified into two types — floppy disks and hard disks. Further, disks that are permanently attached to the unit assembly and cannot be removed by the occasional user are called hard disks. A drive using removable disks is called a floppy disk drive.

(iii) Floppy Disks

NOTES

The disks used with a floppy disk drive are small, removable disks made of plastic, and coated with magnetic recording material. There are two sizes commonly used, with diameters of 5.25 and 3.5 inches.

- The 5.25 inch disk is a 5.25 inch diameter floppy disk. Earlier, such disks recorded data only on one side and were called Single-Sided (SS) disks. Today both the surfaces are used for recording and are called Double-Sided (DS) disks. These are available in two capacities — Double Density (DD), and High Density (HD), where density refers to the number of bits that can be stored per square inch area.

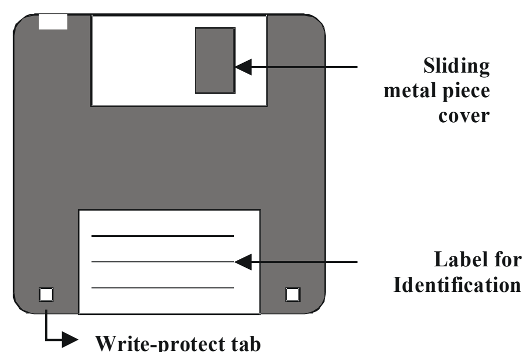


Fig. 2.31 *A 3.5 Inch Floppy Disk*

- The 3.5 inch disk is a disk of 3½ inch diameter. These record data on both sides and are therefore double sided disks. Figure 2.31 displays a 3.5 inch floppy disk. These disks come in three different capacities — double density, high density and very high density. These are smaller and can store more data than can the 5.25 inch disks.

  The storage capacity for any disk can be calculated as:

  Storage capacity = Number of recording surfaces × Number of tracks per surface × Number of sectors per track × Number of bytes per sector

Thus, for a 3.5 inch high density disk which has 80 tracks, 18 sectors/track, and 512 bytes/sector, the disk storage capacity can be calculated as follows:

$$2 \times 80 \times 18 \times 512 = 14,74,560 \text{ bytes or } 1.4 \text{ MB (approximately)}$$

Table 2.1 provides the necessary details and associated storage capacities of various kinds of floppy disks.

Floppy disks were extensively used in personal computers as a medium for distributing software to computer users. Nowadays, CDs or DVDs are used for that purpose.

NOTES

Table 2.1 Storage Capacities of Floppy Disks

Size (diameter in inches)	No. of Recording Surfaces	No. of Tracks	No. of Sectors/ Tracks	No. of Bytes/ Sector	Storage Capacity (approx)
5.25	2	40	9	512	3,68,640 bytes or 360kB
5.25	2	80	15	512	12,28,800 bytes or 1.2 MB
3.5	2	40	18	512	7,37,280 bytes or 720 kB
3.5	2	80	18	512	14,74,560 bytes or 1.4 MB
3.5	2	80	36	512	29,49,120 or 2.8 MB

(ii) Hard Disks

Unlike floppy disks, hard disks are made up of rigid metal. The sizes for the disk platters range between 1 to 14 inches in diameter. Depending on the way they are packaged, hard disks can be categorized as disk packs or Winchester disks.

- **Disk Packs** consist of two or more hard disks mounted on a single central shaft. Because of this, all disks in a disk pack rotate at the same speed. It consists of separate read/write heads for each surface (excluding the upper surface of the topmost disk platter and the lower surface of the bottommost disk platter, as mentioned earlier). Disk packs are removable in the sense that they can be removed and kept offline when not in use (typically stored away in plastic cases) as displayed in Figure 2.32. They have to be mounted on the disk drive before they can be used. Thus different disk packs can be mounted on the same disk drive at different instances, thereby providing virtually unlimited (modular) storage capacity.

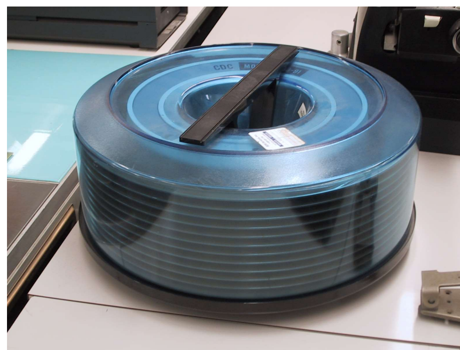


Fig. 2.32 *A Disk Pack*

- ***Winchester Disks*** also consist of two or more hard disk platters mounted on a single central shaft but are of the fixed type. The disk platters are

sealed in a contamination-free container. Due to this fact, all the disk platters, including the upper surface of the topmost disk platter and the lower surface of the bottommost platter, are used for storing data. So, even though Winchester disks have limited storage capacity as opposed to disk packs, they can store larger amounts of data as compared to the same number of disk platters. Figure 2.33 displays a winchester disk.

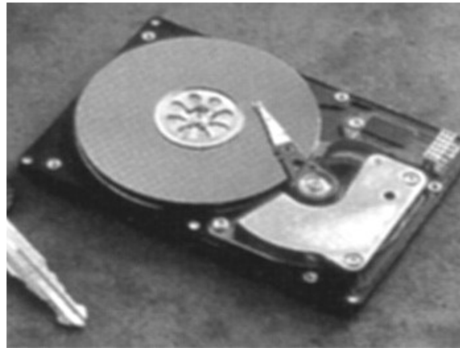


Fig. 2.33 *A Winchester Disk*

Another kind of disk called the zip disk is very common today. This consists of a single hard disk platter encased in a plastic cartridge. Such a disk typically has a capacity of about 100 MB. The zip drive can further be fixed or portable. The fixed zip drive is permanently connected to the computer system while the portable ones can be carried around and connected to any computer system for the duration of its use. In both cases, however, the zip cartridge (the actual storage medium) is portable just like a floppy, albeit with a nearly 100 times larger storage capacity. Figure 2.34 displays zip disks and zip drive.



Fig. 2.34 *Zip Disks and Zip Drive*

5. Optical Storage Devices

(i) Optical Disks

Optical disks are storage devices with huge storage capacities. They are a relatively new storage medium and use laser beam technology for writing and reading of data.

NOTES

Optical disks consist of one large track that starts from the outer edge and spirals inward towards the center (this is unlike the magnetic disk in which tracks are concentric circles on the disk platter). An optical disk is also split into sectors but these are of the same length regardless of their location on the track. Data is therefore packed at maximum density over the disk.

The storage capacity of an optical disk is determined as follows:

Storage Capacity = Number of sectors × Number of bytes per sector

(Note that we do not consider the number of tracks since there is only one track in this case)

Thus, a 5.25 inch optical disk having 3,30,000 sectors and storing 2,352 bytes per sector, will have a storage capacity

3,30,000 × 2352 = 77,61,60,000 bytes or 740 MB (approx.)

The technology used in optical disks uses laser beams to write and read data as opposed to the read /write head used in magnetic disks. Data is recorded by etching microscopic pits (burnt surface) on the disk surface. A high-intensity laser beam is used to etch the pits, while for retrieving data, a low-intensity laser beam is used. Figure 2.35 displays an optical disk formats.



Fig. 2.35 *An Optical Disk and Disk Drive*

There are three optical memory devices that are becoming increasingly popular in various computer applications: CD-ROM, WORM, and Erasable Optical disks. We shall discuss these in the succeeding section.

CD-ROM: The Compact Disk Read-Only Memory(CD-ROM) is a direct extension of the audio CD. It is generally made from a resin called polycarbonate that is coated with aluminium to form a highly reflective surface. The information on a CD-ROM is stored as a series of microscopic pits on the reflective surface (using a high-intensity laser beam). The process of recording information on these disks is known as 'mastering'. This is so-called because this master disk is then used to make a 'die'that is used to make copies.

The information is retrieved from a CD-ROM using a low-powered laser that is generated in an optical disk drive unit. The disk is rotated and the laser beam is aimed at the disk. The intensity of the laser beam changes when it encounters

NOTES

a pit. A photo-sensor detects the change in intensity, hence recognizing the digital signals that are recorded on the surface of the CD-ROM and converts them into electronic signals of 1s and 0s.

As the name suggests, information stored in a CD-ROM can be read only. It cannot be modified in any way. It is therefore useful for applications in which there is a database of information that is useful as it is and does not need changing in any way, for example, a directory such as Yellow Pages. CD-ROMs are very handy in the distribution of large amounts of information to a large number of users. The strengths of CD-ROMs lie in the fact that they provide for:

- Large storage capacity for information/data.
- Fast and inexpensive mass replication.
- Suitable for archival storage since they are removable disks.

The weaknesses of CD-ROMs are:

- They are read-only and cannot be updated.
- Access time in them is greater than that of a magnetic disk.

WORM: The drawbacks of CD-ROM were partially resolved by the introduction of WORM ('Write-Once, Read Many').

In some applications, only a limited copies of compact disks are required to be made. This makes the CD-ROM production economically not viable. This is because manufacturers duplicate CD-ROMs by using expensive duplication equipment. To overcome such a problem, write-once read-many CDs have been developed.

WORM disks allow users to create their own CDs by using a CD-Recordable (CD-R) drive. This can be attached as a peripheral device to the computer system. WORM disks recorded like this can be read by any CD-ROM drive.

Erasable Optical Disk: The erasable optical disk is the latest development in optical disks. Like magnetic disks, the data in the erasable optical disk can be changed repeatedly. Erasable optical disks are also known as rewritable optical disks.

These disks integrate the magnetic and optical disk technologies to enable rewritable storage with the laser-beam technology and so are also called magneto-optical disks. In such systems, a laser beam is used along with a magnetic field to read or write information on a disk that is coated with a magnetic material.

To write, the laser beam is used to heat a specific spot on the magnetic coated material. At this increased temperature, a magnetic field is applied so as to change the polarization of that spot, thereby recording the data that is required. It may be noted here that this process does not cause any physical changes in the disk. Hence, it can be repeated many times. The degree of rotation of the polarized laser beam reflected from the surface is detected to perform reading function. This implies that as the disk spins, the polarized spots pass under the laser beam, and

NOTES

depending on their orientation or alignment, some of them reflect the light while others scatter it. This produces patterns of 'on' and 'off' that are converted into electronic signals of binary 1s and 0s.

The capacity of an erasable disk is very high in comparison to a magnetic disk; for example, a 5.25 inch optical disk can store around 650 MB of data while Winchester disks normally can store a maximum capacity of 320 MB. This is why magneto-optical disks are ideal for multimedia applications that require large storage capacities.

DVD: Known as the Digital Versatile Disk or the Digital Video Disk, it has the same physical dimensions as that of a CD-ROM, but it can hold up to 4.4GB data on a Single layer disk and up to 8.47GB on a dual layer disk with the maximum data transfer of 27MB/s at 20x speeds. The laser used to read/write data on a DVD is much more precise and has a wavelength of 650mm, which is one reason why a DVD can hold more data.

HD-DVD: This is a high density, mostly single-sided, double-layered optical disc that can hold up to 15GB on a single layer and 30GB on a dual layer disc. The read/write speed on an HD-DVD varies between 36 MBPS to 72 MBPS. These were primarily designed for the storage of high definition videos and large volumes of data. The basic look and feel of an HD-DVD drive and disk is the same as that of a CD-ROM and DVD except that it uses a laser of a different wavelength and the microscopic structure of storage on a disk is different.

Blu-Ray Disc: Another high density optical storage media format is gaining increasing popularity these days. Its main uses are high-definition video and data storage. A dual layer Blu-ray Disc can store 50 GB, almost six times the capacity of a double-dual layer DVD (or more than 10 times if single-layer). The data can be read from 36 MBPS to 432 MBPS which is higher than HD-DVD. When used for HD video playback, the video is encoded on it in MPEG-2, AVC, VC-1 (H.264) format, which is the same as HD-DVD.

With the wide variety of storage choices available and constant increase in their number due to changes in technology, it is impossible to say that a single medium of storage can suit the needs of all. Therefore, it is necessary to make the choice of right media based upon factors like capacity, speed, durability, life span and cost of media.

6. Computer Output Microfilm (COM)

COM refers to a process characterized by copying/printing data from media located on personal computers, mini, or mainframe computers onto a microfilm. It comprises a high-speed recorder, which transfers digital data onto a microfilm applying laser technology, and a processor which develops the microfilm once exposed to the light source.

A computer output microfilm device translates information normally held on magnetic tape into miniature images on a microfilm (also called microfiche–'fiche'

pronounced as 'fish'). The device displays information as characters on a CRT screen and then using photographic methods, records the display onto the film. Drawings and images can also be displayed along with narrative text.

A special reader/printer can be subsequently used to view the processed film. The reader operates on a 'back projection' principle displaying one frame at a time on a translucent screen, typically about A4 size. The printer can then be used selectively to produce a hard copy of what is presented on screen.

Figure 2.36 identifies the various steps in COM production.

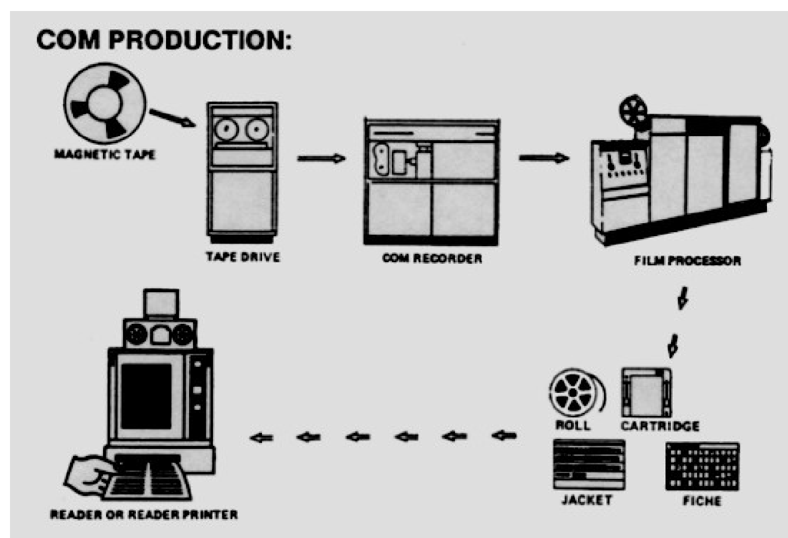


Fig. 2.36 *Steps in a COM Production*

A COM system provides an easy and compact way of recording and storing information, and subsequently retrieving desired pieces of information. It offers various advantages like reduction of paper, reduction in cost (since it is cheaper than most electronic media), improved quality (COM technology provides superior image quality), and electronic record retention/archiving.

COM is best suited for data requiring long-term storage. This is because microfilm is less volatile than magnetic media like disks and tapes. COM stores data in a very compact format. It is to be noted that up to 270 pages can be contained in a single 4 × 6 inch fiche.

Conversion of magnetic tapes to microfilms is cost-effective for closed files. However, in case of highly active data or data requiring regular updating, using microfilms may not be as efficient as retaining the information online. It is useful for data that must be archived for long periods of time and referenced only occasionally, e.g., information that must be archived to comply with legal regulations, information maintained by insurance companies, banks, government agencies, and various other organizations of this type.

NOTES

Measuring Drive Performance

Disk performance can be categorized by the speed at which the data can be read or written. Over the years there have been changes in disk drive interface, rotation speeds, number of heads and cylinders and storage format, all of which have led to a decrease in data access time.

The various types of disks currently available in the market are:

- IDE – Integrated Drive Electronics
- EIDE – Enhanced Integrated Drive Electronics
- SCSI – Small Computer System Interface
- SATA I & II – Serial Advanced Technology Attachment

There are two standard methods for accessing and writing data on a disk — sequential access and random access.

Sequential Access is when you read or write on the disk blocks in sequential or continuous order, that is, one block after another. Examples of where Sequential Access is used in computing or data storage would be the backing up of data onto tape drives or the process of writing data onto CDs and DVDs. Any storage medium based on magnetic tape, VHS, audio cassette etc, are read and written by Sequential Access.

Random Access, as the name suggests, is performed when the hard drive head needs to read/write data from/at various locations on the disk. In this case, the disk heads move rapidly from one place to another and the seek time to access data is increased because it involves mechanical operations. Most of the disk operations performed during routine computer work are random access. This is also the reason why random access time is more important while measuring disk performance than sequential access time. While data is written onto optical media sequentially, data on CDs and DVDs can be read randomly.

For Random Access, the *average seek time* and *average latency time* are added to come up with the total time it takes for the disk to read and write data on it. The average seek time is the time it takes to move the head arm from one position to another, and average latency time is the time it takes for the required data block to come under the head for the read/write operations. The average latency time depends on the Rotations Per Minute (RPM) of the disk, which is the speed at which the magnetic or optical disk rotates.

NOTES

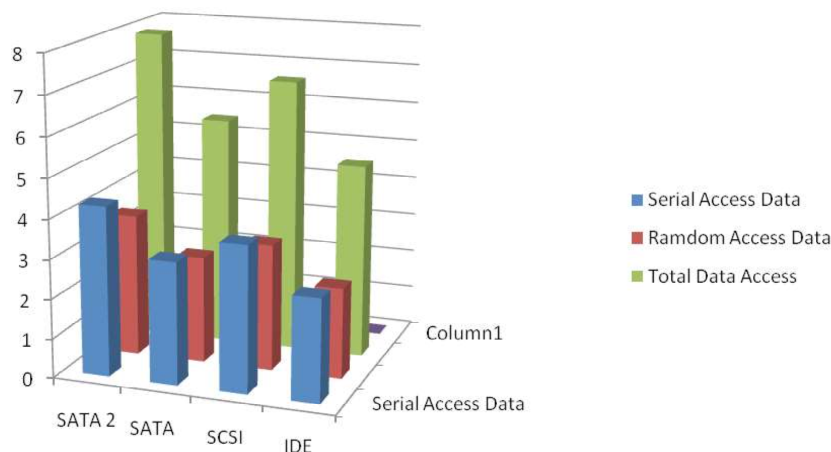


Fig. 2.37 *Data Accessed by Different Disk Drives*

The above bar chart shows the relative amount of data accessed by each type of disk drive.

Check Your Progress

7. What is RAM?
8. What is cache memory?
9. What are disk drives?

2.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The control unit is responsible for the control of various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer.
2. (a) Main power socket: The top part of the rear of the computer, locates the main power cable socket, which supplies power from the electric mains to the computer system. This socket is the part of the main power supply unit of the computer.

   (b) Monitor power socket: The socket that supplies the power from the computer system to the computer monitor is located below the main power cable socket. However, you might not find this socket in all computers and, you can plug in the monitor directly in main power supply.

   (c) PS/2 mouse port: Next you will find a small round green colored port with seven holes and a small logo of the mouse printed next to it. This is where your PS/2 Mouse will be plugged in.

NOTES

3. The most commonly used input devices can be classified into the following categories:
   (a) Keyboard devices (general and special purpose, key-to-tape, key-to-disk, key-to-diskette)
   (b) Point-and-draw devices (mouse, trackball, joystick, light pen, touch screen)
   (c) Scanning devices (optical mark recognition, magnetic ink character recognition, optical barcode reader, digitizer, electronic-card reader)
   (d) Voice recognition devices
   (e) Vision-input devices (webcam, video camera)
4. Scanning devices are input devices used for direct data entry from the source document into the computer system. With the help of the scanner, you can capture your images and documents and convert them into digital formats for easy storage on your computer.
5. Normally, the output is produced on a display unit like a computer monitor or can be printed through a printer on paper. At times, speech outputs and mechanical outputs are also used for some specific applications.
6. Laser printers work on the same printing technology as photocopiers, using static electricity and heat to print with a high quality powder substance known as toner.
7. RAM refers to the mode in which any memory location can be accessed in any order in the same amount of time. Ferrite and semiconductor memories, which usually constitute the primary storage or main memory, are of this nature.
8. Cache memories are small, fast memories placed between the CPU and the main memory. They are faster than the main memory with access times closer to the speed of the CPU.
9. A disk drive is a peripheral device used to store and collect information. It can be removable or fixed, high capacity or low capacity, fast or slow speed, and magnetic or optical.

2.7 SUMMARY

- In order to transfer data to the memory of the computer input devices are used. The Arithmetic Logic Unit (ALU) is responsible for calculations, to which this data from the memory is stored. Once the calculations are done, the data is transferred back to the memory. The memory is responsible for storing data, according to which different functions are carried out. This memory is also known as the main memory or the Immediate Access Store (IAS).

- The control unit is responsible for controlling various computer operations, which involves accepting instructions, interpreting and processing of this information in the correct parts of the computer. It is the main function of the control unit to make sure that the instructions are correctly followed and all operations are done exactly according to the correct instructions at the correct time.
- The primary function of the computer is executing programs. The programs or the set of instructions are stored in the computer's main memory and are executed by the CPU. The CPU processes the set of instructions along with any calculations and comparisons to complete the task. Additionally, the CPU controls and activates various other functions of the computer system. It also activates the peripherals to perform input and output functions.
- In order to store data and instructions, the CPU constitutes many registers, though these are capable of storing very few bytes. All computers need storage space for temporarily storing instruction and data during the execution of the program as the CPU can process data at a speed that is much faster than the speed at which data can be transferred from disks to registers. This could lead to the CPU remaining free most of the time if the data was located in secondary storage including magnetic tapes and disks. The primary or the main memory is the temporary storage located in the computer hardware. Secondary storage or auxiliary memory constitutes devices that can give backup storage, such as magnetic tapes and disks.

  The memory is classified as follows:

  (i) Internal Processor Memory: A small set of high-speed registers placed inside a processor and used for storing temporary data while processing.

  (ii) Primary Storage Memory: The main memory of the computer which communicates directly with the processor. This memory is large in size and fast, but not as fast as the internal memory of the processor. It comprises a couple of integrated chips mounted on a printed circuit board plugged directly on the motherboard. Random Access Memory (RAM) is an example of primary storage memory.

  (iii) Secondary Storage Memory: This stores all the system software and application programs and is basically used for data backups. It is much larger in size and slower than primary storage memory. Hard disk drives, floppy disk drives and flash drives are a few examples of secondary storage memory.

  (iv) Memory Capacity: Capacity, in computers, refers to the number of bytes that it can store in its main memory. This is usually stated in terms of KiloBytes (KB) which is 1024 bytes or MegaBytes (MB)

which is equal to 1024 KB (10,48,576 bytes). The rapidly increasing memory capacity of computer systems has resulted in defining the capacity in terms of GigaBytes (GB) which is 1024 MB (1,07,37,41,824 bytes). Thus a computer system having a memory of 256 MB is capable of storing (256 × 1024 × 1024) 26,84,35,456 bytes or characters.

- The CPU is mostly a microprocessor-based chip located on a single or sometimes a multiple printed circuit boards and is an internal component of the system. It is directly connected to the motherboard; however, the compatibility of the mother board and the CPU depends on the specific series of the latter. Due to the tremendous amount of heat generated by the CPU, it contains a heat sink and a cooling fan.
- The I/O devices that provide a means of communication between the computer and the outside world are known as peripheral devices. This is because they surround the CPU and the memory of a computer system. While input devices are used to enter data from the outside world into the primary storage, output devices are used to provide the processed results from primary storage to users.

  Input devices are used to transfer user data and instructions to the computer.
- Keyboard devices allow input into the computer system by pressing a set of keys mounted on a board, connected to the computer system. Keyboard devices are typically classified as general-purpose keyboards and special-purpose keyboards.
- While working with display-based packages, we usually point to a display area and select an option from the screen (fundamentals of GUI applications). For such cases, the sheer user-friendliness of input devices that can rapidly point to a particular option displayed on screen and support its selection resulted in the advent of various point-and-draw devices.
- Scanning devices are input devices used for direct data entry from the source document into the computer system. With the help of the scanner you can capture your images and documents and convert it into digital formats for easy storage on your computer. The scanner comprises of two major components, the first component lights up the page in order to capture the optical image and the second component converts that captured image into a digital format for viewing and storing it on your computer system.
- Card readers are devices that also allow direct data input into a computer system. The electronic-card reader is connected to a computer system and reads the data encoded on an electronic card and transfers it to the computer system for further processing. Electronic cards are plastic cards with data encoded on them and meant for a specific application.

NOTES

- Voice recognition devices consist of a microphone attached to the computer system. A user speaks into the microphone to input data. The spoken words are then converted into electrical signals (this is in the analog form). A digital-to-analog converter then converts the analog form to digital form (0s and 1s) that can be interpreted by the computer. The digitized version is then matched with the existing pre-created dictionary to perform the necessary action.
- An electromechanical device that accepts data from the computer and translates it into a form that can be understood by the external world is known as an output device. The processed data, stored in the memory of the computer, is sent to an output unit, which then transforms the internal representation of data into a form that can be read by the users.
- A display device is the most essential peripheral of a computer system. Initially, alphanumeric display terminals were used that formed a 7×5 or 9×7 array of dots to display text characters only. As a result of the increasing demand and use of graphics and GUIs, graphic display units were introduced. These graphic display units are based on series of dots known as pixels used to display images. Every single dot displayed on the screen can be addressed uniquely and directly.
- A multimedia projector is an output device which is used to project information from the computer onto a large screen so that it can be viewed by a large group of people.
- A multimedia projector can directly be plugged into a computer system and the information projected on a large screen thereby making it possible to present information to a large audience. The presenter can also use a pointer to emphasize specific areas of interest in the presentation.
- Printers are used for generating output printed on paper. There is a huge range of commercially available printers today (estimated to be 1500 different types).
- Plotters are used to make line illustrations on paper. They are capable of producing charts, drawings, graphics, maps and so on. A plotter is much like a printer but is designed to print graphs instead of alphanumeric characters. Based on the technology used, there are mainly two type of plotters: *pen plotters* and *electrostatic plotters*. Pen plotters, also known as flatbed plotters, draw images with multicolored pens attached to a mechanical arm. Electrostatic plotters, also known as drum plotters, work on the same technology as laser printers.
- Storage Capacity: Represents the size of the memory. It is the amount of data that can be stored in the storage unit. Primary storage units have less storage capacity as compared to secondary storage units. The capacity of

NOTES

internal memory and main memory can be expressed in terms of the number of words or bytes.The capacity of external or secondary storage, on the other hand, is measured in terms of bytes.

- The Static RAM (SRAM) stores binary information using clocked sequential circuits. The stored information remains valid only as long as power is applied to the unit. On the other hand, Dynamic RAM (DRAM) stores binary information in the form of electric charges that are applied to capacitors inside the chip. The stored charge on the capacitors tends to discharge with time and so must be periodically recharged by refreshing the dynamic memory. The Dynamic RAM offers larger storage capacity and reduced power consumption. Therefore, large memories use Dynamic RAM, while static RAM is mainly used for specialized applications.
- RAM was used to refer to random-access memory, but now we use the term read/write memory to distinguish it from Read-Only Memory (since ROM is also random access). RAM is used for storing the bulk of the programs and data that are subject to change, while ROM is used to store programs that are permanently resident in the computer and do not change once the production of the computer is completed.
- Cache memories are small, fast memories placed between the CPU and the main memory. They are faster than the main memory with access times closer to the speed of the CPU. Caches are fast, but very expensive. So, they are used only in small quantities. As an example, caches of size 64K, 128K are normally used in PC-386 and PC-486, which can have 1 to 8 MB of RAM or even more. Cache memories provide fast speed memory retrieval without compromising the size of the memory.
- Magnetic tapes are used for storing files of data that are sequentially accessed or not used very often and are stored off-line. They are typically used as backup storage for archiving of data.
- Magnetic disks are direct-access medium, and so are the most popular online secondary storage devices. Direct-access devices are also called random-access devices because information is literally available at random or in any order. There is direct access to any location on the device.
- Optical disks are storage devices with huge storage capacities. They are a relatively new storage medium and use laser beam technology for writing and reading of data.

  Optical disks consist of one large track that starts from the outer edge and spirals inward towards the center (this is unlike the magnetic disk in which tracks are concentric circles on the disk platter).
- The Compact Disk Read-Only Memory(CD-ROM) is a direct extension of the audio CD. It is generally made from a resin called polycarbonate that is coated with aluminium to form a highly reflective surface. The information

on a CD-ROM is stored as a series of microscopic pits on the reflective surface (using a high-intensity laser beam). The process of recording information on these disks is known as 'mastering'.

- Known as the Digital Versatile Disk or the Digital Video Disk, it has the same physical dimensions as that of a CD-ROM, but it can hold up to 4.4GB data on a Single layer disk and up to 8.47GB on a dual layer disk with the maximum data transfer of 27MB/s at 20x speeds. The laser used to read/write data on a DVD is much more precise and has a wavelength of 650mm, which is one reason why a DVD can hold more data.
- This is a high density, mostly single-sided, double-layered optical disc that can hold up to 15GB on a single layer and 30GB on a dual layer disc. The read/write speed on an HD-DVD varies between 36 MBPS to 72 MBPS. These were primarily designed for the storage of high definition videos and large volumes of data. The basic look and feel of an HD-DVD drive and disk is the same as that of a CD-ROM and DVD except that it uses a laser of a different wavelength and the microscopic structure of storage on a disk is different.
- Disk performance can be categorized by the speed at which the data can be read or written. Over the years there have been changes in disk drive interface, rotation speeds, number of heads and cylinders and storage format, all of which have led to a decrease in data access time.

2.8 KEY WORDS

- **Mouse:** A small input device used to move the cursor on a computer screen to give instructions to the computer and to run programs and applications.
- **Touch pads:** The input devices that are touch sensitive and take user input to manage onscreen pointer and other functions.
- **Light pen:** A small input device used to select and display objects on a screen.
- **Trackball:** A pointing device that is like an inverted mouse.
- **Scanning devices:** Input devices used for direct data entry from the source document into the computer system.
- **Web camera:** A video capturing device attached to the computer system, mostly using a USB port used for video conferencing, video Security, as a control input device.
- **Multimedia projector:** An output device which is used to project information from the computer onto a large screen so that it can be viewed by a large group of people.

- **Magnetic tapes:** These are used for storing files of data that are sequentially accessed or not used very often and are stored off line.
- **Disk drive:** A peripheral device used to store and collect information.

NOTES

2.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are the components of a digital computer?
2. Define motherboard.
3. Write a short note on CPU.
4. Differentiate between input and output devices.
5. State the significance of an electronic-card reader.
6. What do you mean by joystick?
7. State the difference between OMR and OCR.
8. What are the types of printers?
9. Define storages devices along with its types.

Long-Answer Questions

1. Describe the various components of a digital system along with the types of sockets and ports.
2. Briefly describe about the various types computer input devices giving details of each type.
3. Explain briefly about the various types of computer output devices giving specification of each type.
4. Briefly discuss about the types of printers on the basis of their functionalities. Give examples of each type along with the features.
5. Classify different categories of the storage devices giving appropriate examples of each type.
6. Elaborate on the common properties used for characterization of storage units.

2.10 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 3 COMPUTER LANGUAGES

NOTES

Structure

3.0 INTRODUCTION

In computer science, computer languages are systems of communication with a computer. Such languages are used to create computer code or program code, the set of instructions forming a computer program which is executed by the computer. It is one of two components of the software which runs on computer hardware, the other being the data.

Computers can only execute the machine code instructions which are part of their instruction set. Because these instructions are difficult for humans to read, and writing complex programs in machine code or other low level programming languages is a time-consuming task, most programmers write their source code in a high level programming language. This source code is translated into machine code by a compiler or interpreter, so that the computer can execute it to perform its tasks.

A programming language is a formal language comprising a set of instructions that produce various kinds of output. Programming languages are used in computer programming to implement algorithms. Most programming languages consist of instructions for computers. Machine languages were termed as First Generation Languages (1GLs). The Second Generation Languages (2GLs) were termed as Assembly Languages, which were still closely tied to the instruction set architecture of the specific computer. The First High Level Programming Languages, or Third Generation Languages (3GLs), were written in the 1950s. Examples of 3 GL are C, C++, C#, Java, BASIC, PASCAL, FORTRAN, ALGOL, and COBOL. Fourth Generation Languages (4GLs) for programming languages tend to be specialized toward very specific programming domains. 4GL languages may include

support for database management, report generation, mathematical optimization, GUI development, or web development. Examples of 4 GL are SQL, UNIX Shell, Oracle Reports, and R. The Fifth Generation Language (5GL) for programming is referred as the specific programming language that is typically based on problem-solving using constraints given to the program, rather than using an algorithm written by a programmer. Examples of 5GL are OPS5 and Mercury.

In this unit, you will study about the computer languages, types of computer languages, machine language, assembly language and high level languages.

3.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the significance of computer languages
- Explain the different types of computer languages
- Define machine language and assembly language
- Describe the generation of computer languages
- Explain the various computer languages
- Analyse the importance of high level languages
- Elaborate on the advantages and disadvantages of different programming languages

3.2 GENERATION OF COMPUTER LANGUAGES

Computer languages have changed with every changing generation of computers.

The computer languages are systems of communication with a computer. Such languages are used to create computer code or program code, the set of instructions forming a computer program which is executed by the computer. It is one of two components of the software which runs on computer hardware, the other being the data.

Computers can only execute the machine code instructions which are part of their instruction set. Because these instructions are difficult for humans to read, and writing complex programs in machine code or other low level programming languages is a time-consuming task, most programmers write their source code in a high level programming language. This source code is translated into machine code by a compiler or interpreter, so that the computer can execute it to perform its tasks. A compiler produces object code which is usually in machine language, but may also be in an intermediate language which is at a lower level than the source. A runtime system is often used to execute object code by linking it with commonly used libraries. Bytecode is a lower level of source which is designed for more efficient interpretation by interpreters.

NOTES

A programming language is a formal language comprising a set of instructions that produce various kinds of output. Programming languages are used in computer programming to implement algorithms. Most programming languages consist of instructions for computers. There are programmable machines that use a set of specific instructions, rather than general programming languages. Basically, there are two main categories of computer programming languages, namely Low Level Language (LLL) and High Level Language (HLL). Programming languages have been classified into several programming language generations. Historically, this classification or distinction is typically used to specify the increasing power of programming methods.

Following are the five generation of computer programming languages.

1. First Generation Language (1 GL)

The First Generation Language (1GL) for programming are termed as Low Level Languages (LLLs), such as machine level programming languages. The 1GL is a grouping of programming languages that are machine level languages used to program first generation computers. Originally, no translator was used to compile or assemble the first-generation language. The first generation programming instructions were entered through the front panel switches of the computer system. The instructions in 1GL are made of binary numbers, represented by 1s and 0s. This makes the language suitable for the understanding of the machine but far more difficult to interpret and learn by the human programmer.

The main advantage of programming in 1GL is that,

- The code can run very fast and very efficiently, precisely because the instructions are executed directly by the Central Processing Unit (CPU).
- First generation languages are very much adapted to a specific computer and CPU, and code portability is therefore significantly reduced in comparison to higher level languages.
- Modern day programmers still occasionally use machine level code, especially when programming lower level functions of the system, such as drivers, interfaces with firmware and hardware devices.

One of the main disadvantages of programming in a Low Level Language (LLL) is that when an error occurs, the code is not as easy to fix. Modern tools, such as native code compilers are used to produce machine level from a High Level Languages (HLL).

2. Second Generation Language (2 GL)

Second Generation Language (2GL) for programming is a generational way to categorize assembly languages. These are low level assembly languages are used in kernels and hardware drives. The term was coined to provide a distinction from higher level machine independent Third Generation Languages (3GLs) for

programming, such as COBOL and earlier First Generation Languages (1GLs) for programming, such as machine code.

Second Generation Languages (2GLs) for programming have the following properties:

- Lines within a program respond directly to processor commands, essentially acting as a mnemonic device overlaying a first generation programming language.
- The code can be read and written by a programmer.
- To run on a computer it must be converted into a machine readable form, a process called assembly.

The language is specific to a particular processor family and environment.

3. Third Generation Language (3 GL)

A Third Generation Language (3GL) for programming is a generational way to categorize high level computer programming languages.3GLs are much more machine independent and more programmer friendly. This includes features like improved support for aggregate data types, and expressing concepts in a way that favours the programmer, not the computer. A Third Generation Language (3GL) improves over a Second Generation Language (2GL) by having the computer take care of non-essential details. 3GLs are more abstract than previous generations of languages, and thus can be considered High Level Languages (HLLs) than their First Generation Language (1GL) and Second Generation Language (2GL) equivalents. First introduced in the late 1950s, FORTRAN, ALGOL, and COBOL are examples of early 3GLs. Most popular general purpose languages today, such as C, C++, C#, Java, BASIC and PASCAL, are also Third Generation Languages (3GLs).

The main advantage of High Level Languages (HLLs) over Low Level Languages (LLLs) is,

- They are easier to read, write, and maintain.
- Programs written in a High Level Language (HLL) can be easily translated into machine language by a compiler or directly into behaviour by an interpreter.
- These programs could run on different machines so they were machine independent.

Most 3GLs support structured programming while many support Object Oriented Programming (OOP).

4. Fourth Generation Languages (4 GL)

These are non-procedural languages, which suggests that they present the objective but not the procedure to achieve it. The main features of Fourth Generation Languages or 4GLs are as follows:

NOTES

- They constitute simple instructions.
- They are user-friendly, enabling the user to write programs and achieve the desired goal.
- They eliminate the need of a professional programmer for writing programs.

FORTH, was the first Fourth Generation Language (4GLs) and was developed by Charles Morre, an American astronomer, in 1970.

Its main application is in the industrial and scientific control applications. Another example of Fourth Generation Language or 4GL is FOCUS.

5. Fifth Generation Languages (5 GL)

Fifth Generation Languages or 5GLs have developed as a result of research in the area of artificial intelligence. They are, however, early stage.

(i) PROLOG (PROgramming LOGic): A logic programming language. PROLOG a general purpose language often associated with artificial intelligence and computational linguistics. It has a purely logical subset, called Pure PROLOG, as well as a number of extralogical features.

PROLOG was developed in the early 1970s by two French computer scientists, Alain Colmerauer and Philippe Roussel.

Some of other popular languages of 5 GL include:

(ii) GPSS (General Purpose System Simulator): This programming was used for modelling physical and environmental events.

(iii) SNOBOL (String Oriented Symbolic Language): This programming was designed for pattern matching and list processing.

(iv) LOGO (a Version of LISP): This programming language was developed in the 1960s to help children learn simple programs an computers.

(v) PILOT (Programmed Instruction Learning, Or Testing): This programming language was used for writing instructional software.

3.3 COMPUTER LANGUAGES

A computer language essentially implies a language that is understandable to the computer. It is the computer's native language. Computer languages serve the same purpose as human languages. They are a means of communication. Let us understand the similarities and differences between computer languages and human languages.

Languages that the speak daily, such as English, Hindi, French or German are known as material or human languages. It constitutes words and rules known as lexicon and syntax, respectively. These words are joined to make meaningful phrases according to the rules of the syntax. A computer language also consists of lexicon and syntax, i.e., characters, symbols and rules of usage that allow the user to communicate with the computer.

NOTES

The primary difference between a natural language and computer language is that natural languages have a large set of words (vocabulary) to choose from while computer languages have a limited or restricted set of words. Thus, fewer words but more rules characterize a computer language.

All problems to be solved by the computer needs to be broken down into discrete logical steps before the computer can execute them. The process of writing such instructions in a computer or programming language is called programming or coding.

Since as computer hardware has improved over the years, programming languages have also moved from machine oriented languages (that used strings of binary 0s and 1s) to problem-oriented languages (that use common English terms). All computer languages can, however, be classified under the following categories:

- Machine Language (First Generation Language or 1GL)
- Assembly Language (Second Generation Language or 2GL)
- High-Level Language (Third Generation Language or 3GL)

Check Your Progress

1. What is a computer language?
2. What language can a computer understand?
3. Classify different categories of computer languages.
4. State the properties of second level computer language.

3.4 CLASSIFICATION OF COMPUTER LANGUAGES

Computer languages are classified as follows:

3.4.1 Machine Language

The computer can understand only a binary based language. This is a combination of 0s and 1s. Instructions written using sequences of 0s and 1s constitute a are known as machine language. First-generation computers used programs written in machine language.

A major drawback of machine language is that it is highly complex and difficult to use. Also, it consumes a lot of time and requires a substantial effort on the part of the programmer. Thousands of machine language instructions are needed to carry out simple tasks, such as listing a few addresses for mails. Any instruction in machine language is divided into two components:

(i) **Command:** Also called the 'Operation Code' or opcode including addition, multiplication, etc.

NOTES

(ii) **Operand:** Refers to the address of the date on which the function has to be performed.

A general machine language instruction is presented as follows:

OP Code	Operand
001	010001110

The number of operands varies with each computer and is therefore computer dependent.

It can be concluded that in order to develop computer programs in machine language, the programmer will be required to remember a lot of operation codes and addresses of the data items based on the storage location and also information regarding the internal structure of the computer. Thus, using machine language can be highly complicated and liable to errors. Identifying these errors and introducing changes had become increasingly difficult leading programmers to seek better options.

3.4.2 Assembly Language

The development of assembly language marked the beginning of the evolution of programming languages. In assembly language mnemonics (symbolic codes) were used to present operation codes as well as strings of characters to represent addresses. Instructions in assembly language may appear as follows:

Opration	Opration address
READ	M
ADD	L

Certain important facts about assembly language are as follows:

- Assembly language was designed to replace each machine code by an understandable mnemonic and each address with a simple alphanumeric string. It was matched to the processor structure of a particular computer and was therefore (once again) machine dependent. This meant that programs written for a particular computer model could not be executed on another one. In other words, an assembly language program lacked portability.
- A program written in assembly language needs to be translated into machine language before the computer can execute it. This is done by a special program called 'Assembler' which takes every assembly language program and translates it into its equivalent machine code.
- The assembly language program is known as the source program, while the equivalent machine language program is known as the object program. It may be useful to know that the assembler is a system program supplied by the computer manufacturer. Second-generation computers used assembly language.

NOTES

- The lack of portability of programs (written using machine or assembly languages) between various computer systems led to the development of high level languages. Since they allowed a programmer to overlook a lot of low level particulars of the hardware of the computer system, they were called high level language programs.
- It was obvious that if the syntax, mnemonics and rule and regulations of the programming language were closer to the natural language, it would be easier for the programmer to program and the lesser the possibility of introducing errors (or bugs) into the program. Hence, third-generation languages, which were algorithmic and procedural, came into being in the mid-1950s. They were designed to solve a particular problem. They contained commands that are particularly suited to one type of application. For example, a number of languages were designed to process scientific or mathematical problems. Others emphasized on commercial applications. These languages varied very little between different computer systems, unlike machine or symbolic languages. But a complier or an interpreter program was required to translate these machine codes. Once again, the high level program is called the source code while its equivalent machine language program is referred to as the object code.
- Easy-to-learn feature, machine independence, easier maintenance and portability contributed to the popularity of high level languages. Slow program execution was the main disadvantage since programs needed to be converted into machine language (by an interpreter or a compiler) before they could be executed.

A program written in assembly language needs to be translated into machine language before the computer can execute it. This is done by a special program called 'Assembler' which takes every assembly language program and translates it into its equivalent machine code. The assembly language program is known as the source program, while the equivalent machine language program is known as the object program. It may be useful to know that the assembler is a system program supplied by the computer manufacturer. Second-generation computers used assembly language.

The lack of portability of programs (written using machine or assembly languages) between various computer systems led to the development of high level languages. Since they allowed a programmer to overlook a lot of low level particulars of the hardware of the computer system, they were called high level language programs. And it was obvious that if the syntax, mnemonics and rule and regulations of the programming language were closer to the natural language, it would be easier for the programmer to program and the lesser the possibility of introducing errors (or bugs) into the program. Hence, third-generation languages, which were algorithmic and procedural, came into being in the mid-1950s. They were designed to solve a particular problem. They contained commands that are

NOTES

particularly suited to one type of application, for example, a number of languages were designed to process scientific or mathematical problems. Others emphasized on commercial applications. These languages varied very little between different computer systems, unlike machine or symbolic languages. But a complier or an interpreter program was required to translate these machine codes. Once again, the high level program is called the source code while its equivalent machine language program is referred to as the object code.

Easy-to-learn features, machine independence, easier maintenance and portability contributed to the popularity of High Level Languages or HLLs. Slow program execution was the main disadvantage since programs needed to be converted into machine language (by an interpreter or a compiler) before they could be executed.

Check Your Progress

5. Define the two components required for any instruction in machine language?
6. What is an assembler?
7. List few high level languages.

3.4.3 High Level Languages (HLLs)

Some of the High Level Languages or HLLs have been discussed as follows:

1. FORTRAN

FORTAN (FORmula TRANslation) was the first high -level language developed by John Backus at IBM in 1956.

FORTRAN has a number of versions with FORTRAN IV being one of the earlier popular versions. In 1977, the American National Standards Institute (ANSI) published standards for FORTRAN with a view to standardizing the form of the language used by manufacturers. This standardized version is called FORTRAN 77.

2. COBOL

COBOL (Common Business Oriented Language), the first language used for commercial applications, was developed under the leadership of Grace Hopper, a US Navy programmer, with a group of computer manufacturers and users in 1959. The maintenance and further growth of the language was handed over to a group called CODASYL (COnference on Data SYstems Languages).

It is written using statements that resemble simple English and can be understood easily; for example, to add two numbers (stored in variables A and B), a simple statement in COBOL would be: ADD A TO B GIVING C.

COBOL was standardized by ANSI in 1968 and in 1974. COBOL became the most widely used programming language for business and data processing applications.

3. BASIC

BASIC (Beginner's All-Purpose Symbolic Instruction Code) was developed as a teaching tool for undergraduate students in 1966 by John Kemeny and Thomas Kurtz, two professors at Dartmouth College. Eventually BASIC was used as the main language amongst the personal computer users.

NOTES

A minimum version of BASIC was standardized by ANSI and is so simple that it has been incorporated in every subsequent version of BASIC. Some versions of BASIC include MBASIC (Microsoft BASIC), and CBASIC (Compiler based BASIC).

One of the newer versions of BASIC, commonly known as Visual Basic, has also evolved from the original BASIC language. It contains various statements and functions that can be used to create applications for a Windows or GUI environment.

4. PASCAL

PASCAL was designed by Nicholas Wirth, a Swiss professor, in 1971. It was developed as a more structured language used for teaching which Wirth named after the French mathematician Blaise Pascal, who also designed the first successful mechanical calculator. His primary aim was to provide a language that supported beginners learning good problem solving and programming techniques.

In addition to manipulation of numbers, PASCAL supports manipulation of vectors, matrices, strings of characters, records, files and lists, thereby supporting non-numeric programming. Hence, it has proved to be an attractive language for professional computer scientists.

PASCAL has been standardized by ISO (International Standards Organization) and ANSI.

5. PL/1

(Programming Language 1) was developed by IBM in the 1960s and was the first language that was attempted to be used for a variety of applications rather than one particular area like business or science or artificial Intelligence.

6. LISP

(LISP Processing) was developed in the early 1950s but was implemented in the 1959 by John McCarthy at the Massachusetts Institute of Technology. It became a standard language with the artificial intelligence community and was a program that could easily handle recursive.

7. C

This language was developed by Dennis Ritchie of Bell Laboratories in order to implement the operating system UNIX.

NOTES

8. C ++

This language was developed by the Bjarne Stroustrup of Bell Laboratories by enhancing C. C++ is also used to write procedural programs like C but the reason for its increased popularity is perhaps because of its capability to handle the rigours of object-oriented programming. C and C++ are the most extensively used general-purpose languages amongst programming experts.

9. JAVA

Java is again an object-oriented language like the C++ but is a simplified version with extra features. It is less prone to programming errors. It was developed for writing programs that could be safely and easily executed through the Internet. It is free from any kind of common virus threats. It is basically a network-oriented language that can develop website pages with enhanced multimedia features using small java programs known as java applets, Java is a secure to use over the Internet and is a platform independent language.

3.5 MERITS AND DEMERITS OF VARIOUS PROGRAMMING LANGUAGES

Programming languages form an important part of a programmer's tools and are required to create applications in order to enable communication between a user and a computer. The basic aim of programming languages is to provide technical answers to real-life complications, such as remote sensing programs, weather forecasting software, etc. The logic development increases due to program statements, in order to arrive at solutions of complex equations, subroutines, etc. Thus, computer languages find solutions aided by these languages, for instance 'C' language can be used for solving quadratic equations.

Programming languages aid in arriving at quick solutions, using logic and turning them into computer syntax according to the selected language. Programming languages convert the logic into codes, which makes it easy to debug errors. Different programming languages are as follows:

1. C

C allows input and output control and makes it easy to access hardware devices and to do 'bit twiddling', i.e., manipulating individual bits in hardware registers. C is a very powerful language that is used for optimization and efficient faster code execution. It is the first choice of embedded technology development. Table 3.1 explains the merits and demerits of C language.

NOTES

Table 3.1 Merits and Demerits of the C Language

Merits	Demerits
C supports pointers that basically refer to records, the address or location of function or object in a memory.	It has no runtime checking and is a case-sensitive language.
It is used especially in game programming, and it is faster than C++.	There is no strict checking in C code, for example, you can pass an integer value for the float data type value.
It is a structured language, and it extensively uses pointers for memory, array, structures and functions.	If the code extends, it is very difficult to fix the bugs.

2. JavaScript

JavaScript is a scripting language used in client-side script programming on the web. It makes dynamic updates to a HTML document in a browser. It interacts between browsers and users. Table 3.2 explains the merits and demerits of the JavaScript language.

Table 3.2 Merits and Demerits of the JavaScript Language

Merits	Demerits
It runs fast because it is a client-side language.	All browsers do not recognize JavaScript. Some JavaScript versions limit the old browsers.
The JavaScript code runs fast in a browser and makes quick connection with a server.	It depends on the compatibility of a browser. Different JavaScript versions are available for different browsers. For example, Netscape browser supports JavaScript 1.5, Microsoft supports Jscript and Internet Explorer 6.0 supports Jscript 5.6.
To write the JavaScript code, no extra tools are required. Any HTML editor or notepad support it.	It is limited to access and manipulate with HTML code.

3. Common Business Oriented Language (COBOL)

COBOL applications run in critical areas of business. The syntax of COBOL is based on English clauses, such as verbs, sentences, clauses, sections, divisions, etc. In fact, finance and insurance data is processed with COBOL. Table 2.3 explains the merits and demerits of the COBOL language.

Table 3.3 Merits and Demerits of the COBOL Language

Merits	Demerits
COBOL is a robust, understandable, readable and self-documenting language.	It has a rigid hierarchical structure where items are restricted to the environment division.
It is a simple language having no pointers, no user defined functions, and functions are used in a limited scope.	In COBOL coding, programmers cannot write code for mobile devices or screen readers.
It is portable, scalable and maintainable, which reduces maintenance cost. It can work with new peripheral devices.	This language cannot be used for developing web applications.

NOTES

4. ActionScript

ActionScript is the programming language of Adobe Flash platform. It is an object-oriented programming language based on European Computer Manufacturers Associations (ECMA) script. The applications are used to play the browser.

Table 3.4 *Merits and Demerits of the ActionScript Language*

Merits	Demerits
It is best used for animating movies.	It is not suitable for non-web application.
It provides interactive and visual effects of the web browser.	It is not well versed with low-level programming.
It is widely and popularly used on the index.html for commercial websites.	It mainly resides inside the HTML code.

5. C++

C++ is also an Object Oriented Language (OO) that uses virtual functions, multiple inheritance, exception handling, polymorphism, etc. Its object-oriented structure allows the code to be reused that cuts down the development time. C++ provides comprehensive coverage of abstract classes as interfaces, regular error handling, standard strings, I/O streams, etc. Table 3.5 explains the merits and demerits of the C++ language.

Table 3.5 *Merits and Demerits of the C++ Language*

Merits	Demerits
This language works on graphic user interface programming on a computer.	It is complex for very large high-level programs and difficult to debug for web applications.
It is well-suited language to write codes for OS drivers that are able to develop platform-dependent applications.	A library set is used with this programming language that restricts a single platform or OS.
It provides various new operators, such as new, scope resolution operator (::).	Program execution is a little bit slower than C language.

6. C#

C# is a portable programming language and is used in Microsoft.NET framework. It is used to create applications and the output is displayed on a black window referred to as black window or DOS window. Table 3.6 explains the merits and demerits of the C# language.

Table 3.6 *Merits and Demerits of the C# Language*

Merits	Demerits
This language is used for web application under the .NET platform.	It is restricted only for Microsoft platform.
The cost of maintenance is much lower than C++. All features of WPF, such as base class libraries, ADO.NET, etc., are available.	Windows Presentation Foundation (WPF) applications run a little bit slow.
C# is safer to run because OS always checks or verifies malicious coding.	It is always slow to run because after launching the programs animation effects come smooth.

7. Visual BASIC

Visual BASIC (VB) creates educational, financial and accounting applications and is supported by the latest Microsoft OS. Windows Vista supports VB5.0 version. Table 3.7 explains the merits and demerits of the VB language.

Table 3.7 *Merits and Demerits of the VB Language*

Merits	Demerits
Online help is available for coding.	Applications developed in VB can work only on Windows OS.
It provides rapid application environment that easily facilitates drag and drop components.	More memory space is required to work with and install VB, because it is based on GUI tools that contain graphical components.
The syntax is simpler than other high-level languages.	It is not a suitable language to develop codes for video games applications.

8. Java

Java technology has become a complete software ecosystem that represents different values to different types of consumers and business users. Table 3.8 explains the merits and demerits of the Java language.

Table 3.8 *Merits and Demerits of the Java Language*

Merits	Demerits
Java is a simple, object oriented, secure, robust, multi-threaded and platform-independent language. Java programs are interpreted and then compiled into Java Virtual Machine (JVM) code called bytecode.	Java uses swing toolkit that is very different from native applications.
It is a single-paradigm language and static imports are available. It supports three Java platform editions, such as Java technology in mobile devices, in desktops and medium to large business.	It is slow and consumes more memory.
Java language is designed to make distributed computing easy that provides networking capability for system unit via programs.	There are some hardware and software requirements in Java language, for example, hardware requirements (512 MB of RAM, 1000 MHz of processor speed, Pentium III) and software requirements (Microsoft Windows OS, Java Runtime Environment and Internet Explorer).

9. Java 2 Platform Enterprise Edition (J2EE)

J2EE platform uses distributed multitiered (client, web and business tiers) application model. Table 3.9 explains the merits and demerits of the J2EE language.

Table 3.9 Merits and Demerits of the J2EE Language

NOTES

Merits	Demerits
This language is capable of implementing 3D graphics games.	Many errors occur only at runtime.
It is a web framework that is easy to develop and debug. It gives complete architecture for developing, such as session tracking and managing transaction.	Several steps are required, for example, compiling, packaging, deploying and running the application to run a complete program.
The applet distribution and updates are simplified. It is a separator of application specific code.	It is a very complex technology. A simple program includes many interfaces, bean classes and deployment descriptors.

10. Python

Python is an object-oriented language that dynamically develops software applications. It provides many tools to write socket programming, for example, server-client programs. It is a maintainable code and easy to learn. Table 3.10 explains the merits and demerits of Phython

Table 3.10 Merits and Demerits of the Python Language

Merits	Demerits
Python provides different types of variables in an array but the types of variables are not specified. The % operand allows quick scripting workflow.	This language is executed by an interpreter, which causes to slow the programs execution.
This language provides the readability and scalability of codes for large projects.	Global namespaces of Python modules are not always deallocated.
Python, like COBOL, is a Universal Turing Machine. Python is not only a good language for CGI scripting but also for other tasks, especially in middleware environments.	The MEL commands are first echoed and then translated.

11. Ruby on Rails

Ruby is an interpreted scripting object-oriented programming language where Rails supports a full-stack and open source web framework. Table 3.11 explains the merits and demerits of the Ruby On Rails language.

Table 3.11 Merits and Demerits of the Ruby on Rails Language

Merits	Demerits
It is a powerful language and has rich libraries.	Implementations of threads are not native.
It is easy to install. It is more object oriented than Python.	It supports only single inheritance.
It comes from the SmallTalk 'Everything on Objects'. It supports strong pattern matching techniques.	It works with former object hierarchy.

NOTES

12. Hypertext Preprocessor (PHP)

PHP is basically the open source of server-side scripting language. PHP runs on Windows OS and UNIX OS. Apache server is needed to execute the program that provides local host explorer services. It mainly contains index.html file inside the folder to go with web pages. The cookies (small data file) are set within the program to get a requested web page. Table 3.12 explains the merits and demerits of the PHP language.

Table 3.12 *Merits and Demerits of the PHP Language*

Merits	Demerits
It is a script language for dynamic web page design.	The cascading style sheet (CSS) can be maintained in HTML. You can add JavaScript code in PHP to interact with MySQL database.
This language is easy to link database, such as MySQL, Oracle and is able to generate HTTP headers.	If programmers modify the source code the previous output stays on the buffer, so every time they have to refresh the web page by pressing [F5] function key.
It is an efficient competitor of Microsoft's ASP. It is portable and embedded with HTML. Error handling is adequately available now in PHP5, allowing you to use try/catch exception handling and overriding the default error handler to invoke your own methods to tracing back the stack.	Each time the same computer requests a page with a browser, it will send the cookie with PHP that takes more time to execute successfully.

Check Your Progress

8. State the use of ActionScript.
9. What are the merits of PHP language?

3.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A computer language also consists of lexicon and syntax, i.e., characters, symbols and rules of usage that allow the user to communicate with the computer.
2. The computer can understand only a binary based language. This is a combination of 0s and 1s. Instructions written using sequences of 0s and 1s constitute a are known as machine language.
3. All computer languages can, however, be classified under the following categories:
   - Machine Language (First Generation Language or 1GL)
   - Assembly Language (Second Generation Language or 2GL)
   - High-Level Language (Third Generation Language or 3GL)

NOTES

5. Any instruction in machine language is divided into two components:
   (i) Command: Also called the 'Operation Code' or opcode including addition, multiplication, etc.
   (ii) Operand: Refers to the address of the date on which the function has to be performed.

   A general machine language instruction is presented as follows:

OP Code	Operand
001	010001110

6. A program written in assembly language needs to be translated into machine language before the computer can execute it. This is done by a special program called 'Assembler' which takes every assembly language program and translates it into its equivalent machine code. The assembly language program is known as the source program, while the equivalent machine language program is known as the object program. It may be useful to know that the assembler is a system program supplied by the computer manufacturer. Second-generation computers used assembly language.
8. ActionScript is the programming language of Adobe Flash platform. It is an object-oriented programming language based on European Computer Manufacturers Associations (ECMA) script. The applications are used to play the browser. The merits of ActionScript programming languages are:
   - It is best used for animating movies.
   - It provides interactive and visual effects of the web browser.
   - It is widely and popularly used on the index.html for commercial websites.
9. Following are the merits of PHP language:
   - It is a script language for dynamic web page design.
   - This language is easy to link database, such as MySQL, Oracle and is able to generate HTTP headers.
   - It is an efficient competitor of Microsoft's ASP. It is portable and embedded with HTML. Error handling is adequately available now in PHP5, allowing you to use try/catch exception handling and overriding the default error handler to invoke your own methods to tracing back the stack.

3.7 SUMMARY

- The computer languages are systems of communication with a computer. Such languages are used to create computer code or program code, the set of instructions forming a computer program which is executed by the computer. It is one of two components of the software which runs on computer hardware, the other being the data.

NOTES

- A programming language is a formal language comprising a set of instructions that produce various kinds of output. Programming languages are used in computer programming to implement algorithms. Most programming languages consist of instructions for computers. There are programmable machines that use a set of specific instructions, rather than general programming languages.
- The First Generation Language (1GL) for programming are termed as Low Level Languages (LLLs), such as machine level programming languages. The 1GL is a grouping of programming languages that are machine level languages used to program first generation computers. Originally, no translator was used to compile or assemble the first-generation language. The first generation programming instructions were entered through the front panel switches of the computer system. The instructions in 1GL are made of binary numbers, represented by 1s and 0s.
- Second Generation Language (2GL) for programming is a generational way to categorize assembly languages. These are low level assembly languages are used in kernels and hardware drives. The term was coined to provide a distinction from higher level machine independent Third Generation Languages (3GLs) for programming, such as COBOL and earlier First Generation Languages (1GLs) for programming, such as machine code.
- A Third Generation Language (3GL) for programming is a generational way to categorize high level computer programming languages.3GLs are much more machine independent and more programmer friendly. This includes features like improved support for aggregate data types, and expressing concepts in a way that favours the programmer, not the computer.
- The main features of Fourth Generation Languages or 4GLs are as follows:
  - o They constitute simple instructions.
  - o They are user-friendly, enabling the user to write programs and achieve the desired goal.
  - o They eliminate the need of a professional programmer for writing programs.
- Some of other popular languages of 5 GL include:

  (ii) GPSS (General Purpose System Simulator): This programming was used for modelling physical and environmental events.

  (iii) SNOBOL (String Oriented Symbolic Language): This programming was designed for pattern matching and list processing.

  (iv) LOGO (a Version of LISP): This programming language was developed in the 1960s to help children learn simple programs an computers.

  (v) PILOT (Programmed Instruction Learning, Or Testing): This programming language was used for writing instructional software.

NOTES

- A computer language essentially implies a language that is understandable to the computer. It is the computer's native language. Computer languages serve the same purpose as human languages. They are a means of communication.
- All problems to be solved by the computer needs to be broken down into discrete logical steps before the computer can execute them. The process of writing such instructions in a computer or programming language is called programming or coding.
- All computer languages can, however, be classified under the following categories:
  - o Machine Language (First Generation Language or 1GL)
  - o Assembly Language (Second Generation Language or 2GL)
  - o High-Level Language (Third Generation Language or 3GL)
- Any instruction in machine language is divided into two components:
  - (i) **Command:** Also called the 'Operation Code' or opcode including addition, multiplication, etc.
  - (ii) **Operand:** Refers to the address of the date on which the function has to be performed.
- A program written in assembly language needs to be translated into machine language before the computer can execute it. This is done by a special program called 'Assembler' which takes every assembly language program and translates it into its equivalent machine code.
- A program written in assembly language needs to be translated into machine language before the computer can execute it. This is done by a special program called 'Assembler' which takes every assembly language program and translates it into its equivalent machine code. The assembly language program is known as the source program, while the equivalent machine language program is known as the object program. It may be useful to know that the assembler is a system program supplied by the computer manufacturer. Second-generation computers used assembly language.
- FORTAN (FORmula TRANslation) was the first high -level language developed by John Backus at IBM in 1956.
- COBOL (Common Business Oriented Language), the first language used for commercial applications, was developed under the leadership of Grace Hopper, a US Navy programmer, with a group of computer manufacturers and users in 1959. The maintenance and further growth of the language was handed over to a group called CODASYL (COnference on Data SYstems Languages).
- BASIC (Beginner's All-Purpose Symbolic Instruction Code) was developed as a teaching tool for undergraduate students in 1966 by John Kemeny and

Thomas Kurtz, two professors at Dartmouth College. Eventually BASIC was used as the main language amongst the personal computer users.

- PASCAL was designed by Nicholas Wirth, a Swiss professor, in 1971. It was developed as a more structured language used for teaching which Wirth named after the French mathematician Blaise Pascal, who also designed the first successful mechanical calculator. His primary aim was to provide a language that supported beginners learning good problem solving and programming techniques.
- (Programming Language 1) was developed by IBM in the 1960s and was the first language that was attempted to be used for a variety of applications rather than one particular area like business or science or artificial Intelligence.
- (LISP Processing) was developed in the early 1950s but was implemented in the 1959 by John McCarthy at the Massachusetts Institute of Technology. It became a standard language with the artificial intelligence community and was a program that could easily handle recursive.
- This language was developed by Dennis Ritchie of Bell Laboratories in order to implement the operating system UNIX.
- This language was developed by the Bjarne Stroustrup of Bell Laboratories by enhancing C. C++ is also used to write procedural programs like C but the reason for its increased popularity is perhaps because of its capability to handle the rigours of object-oriented programming. C and C++ are the most extensively used general-purpose languages amongst programming experts.
- Java is again an object-oriented language like the C++ but is a simplified version with extra features. It is less prone to programming errors. It was developed for writing programs that could be safely and easily executed through the Internet.
- Programming languages form an important part of a programmer's tools and are required to create applications in order to enable communication between a user and a computer. The basic aim of programming languages is to provide technical answers to real-life complications, such as remote sensing programs, weather forecasting software, etc. The logic development increases due to program statements, in order to arrive at solutions of complex equations, subroutines, etc. Thus, computer languages find solutions aided by these languages, for instance 'C' language can be used for solving quadratic equations.
- C allows input and output control and makes it easy to access hardware devices and to do 'bit twiddling', i.e., manipulating individual bits in hardware registers. C is a very powerful language that is used for optimization and efficient faster code execution. It is the first choice of embedded technology development.

NOTES

NOTES

- JavaScript is a scripting language used in client-side script programming on the web. It makes dynamic updates to a HTML document in a browser. It interacts between browsers and users.
- COBOL applications run in critical areas of business. The syntax of COBOL is based on English clauses, such as verbs, sentences, clauses, sections, divisions, etc. In fact, finance and insurance data is processed with COBOL.
- ActionScript is the programming language of Adobe Flash platform. It is an object-oriented programming language based on European Computer Manufacturers Associations (ECMA) script. The applications are used to play the browser.
- C++ is also an Object Oriented Language (OO) that uses virtual functions, multiple inheritance, exception handling, polymorphism, etc. Its object-oriented structure allows the code to be reused that cuts down the development time. C++ provides comprehensive coverage of abstract classes as interfaces, regular error handling, standard strings, I/O streams, etc.
- C# is a portable programming language and is used in Microsoft.NET framework. It is used to create applications and the output is displayed on a black window referred to as black window or DOS window.
- Visual BASIC (VB) creates educational, financial and accounting applications and is supported by the latest Microsoft OS. Windows Vista supports VB5.0 version.
- Java technology has become a complete software ecosystem that represents different values to different types of consumers and business users.
- J2EE platform uses distributed multitiered (client, web and business tiers) application model.
- Python is an object-oriented language that dynamically develops software applications. It provides many tools to write socket programming, for example, server-client programs. It is a maintainable code and easy to learn.
- Ruby is an interpreted scripting object-oriented programming language where Rails supports a full-stack and open source web framework.
- PHP is basically the open source of server-side scripting language. PHP runs on Windows OS and UNIX OS. Apache server is needed to execute the program that provides local host explorer services. It mainly contains index.html file inside the folder to go with web pages. The cookies (small data file) are set within the program to get a requested web page.

3.8 KEY WORDS

- **FORTRAN (FORmula TRANslation):** FORTRAN was the first high level language developed by John Backus at IBM in 1956.

NOTES

- **COBOL (Common Business Oriented Language):** The COBOL was the first language used for commercial applications, was developed under the leadership of Grace Hopper.
- **LISP (LISP Processing):** LISP was developed in the early 1950s but was implemented in the 1959 by John McCarthy at the Massachusetts Institute of Technology.
- **PL/1 (Programming Language 1):** PL1 was developed by IBM in the 1960s and was the first language that was attempted to be used for a variety of applications rather than one particular area like business or science or artificial Intelligence.
- **C++:** The programming language C++ was developed by the Bjarne Stroustrup of Bell Laboratories by enhancing C.
- **SNOBOL (String Oriented Symbolic Language):** The programming language SNOBOL was designed for pattern matching and list processing.
- **GPSS (General Purpose System Simulator):** The programming language GPSS was used for modelling physical and environmental events.
- **LOGO (a version of LISP):** The programming language LOGO was developed in the 1960s to help children learn simple programming applications on computers.
- **PILOT (Programmed Instruction Learning, Or Testing):** The programming language PILOT was typically used for writing instructional software.

3.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the importance of computer languages.
2. What is meant by 4GL?
3. State the properties of third level programming language.
4. Which language can be directly understood by the CPU?
5. What are the difference between C and C++?
6. What are the merits and demerits of J2EE?
7. State the definition of LISP.

Long-Answer Questions

1. Explain briefly the classification system of the computer languages.
2. Briefly describe the evolution of the computer languages giving appropriate examples.

NOTES

3. Describe the characteristic features of C, C++, C#, JavaScript, Java, Visual BASICS, J2EE, Python and PHP.
4. Elaborate on the significant properties of COBOL giving its merits and demerits.
5. Briefly discuss the basic properties of machine language, assembly language and high level languages giving appropriate examples of each type.

3.10 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 4 OPERATING SYSTEM AND NETWARE

NOTES

Structure

4.0 INTRODUCTION

An Operating System (OS) is system software that manages computer hardware, software resources, and provides common services for computer programs. Time-sharing operating systems schedule tasks for efficient use of the system and may also include accounting software for cost allocation of processor time, mass storage, printing, and other resources. For hardware functions, such as input and output and memory allocation, the operating system acts as an intermediary between programs and the computer hardware, although the application code is usually executed directly by the hardware and frequently makes system calls to an OS function or is interrupted by it. Operating systems are found on many devices that contain a computer – from cellular phones and video game consoles to web servers and supercomputers. NetWare is a discontinued computer network operating system developed by Novell, Inc. It initially used cooperative multitasking to run various services on a personal computer, using the IPX network protocol.

MS-DOS, acronym for Microsoft Disk Operating System, is an operating system for x86-based personal computers mostly developed by Microsoft. Collectively, MS-DOS, it's rebranding as IBM PC DOS, and some operating systems attempting to be compatible with MS-DOS, are sometimes referred to as

NOTES

'DOS', which is also the generic acronym for disk operating system. MS-DOS was the main operating system for IBM PC compatible personal computers during the 1980s, from which point it was gradually superseded by operating systems offering a Graphical User Interface (GUI), in various generations of the graphical Microsoft Windows operating system.

UNIX was originally written in assembly language. Ken Thompson wrote B, mainly based on BCPL, based on his experience in the MULTICS project. B was replaced by C, and UNIX, rewritten in C, developed into a large, complex family of inter-related operating systems which have been influential in every modern operating system.

Microsoft introduced an operating environment named Windows on November 20, 1985, as a graphical operating system shell for MS-DOS in response to the growing interest in Graphical User Interfaces (GUIs). As of February 2020, the most recent version of Windows for PCs, tablets and embedded devices is Windows 10. Microsoft Windows, commonly referred to as Windows, is a group of several proprietary graphical operating system families, all of which are developed and marketed by Microsoft. Active Microsoft Windows families include Windows NT and Windows IoT (Internet of Things); these may encompass subfamilies, such as Windows Server or Windows Embedded Compact (Windows CE). Defunct Microsoft Windows families include Windows 9x, Windows Mobile and Windows Phone.

Other specialized classes of operating systems, such as embedded and real-time systems, exist for many applications. Embedded operating systems are designed to be used in embedded computer systems. They are designed to operate on small machines with less autonomy, such as PDAs. They are very compact and extremely efficient by design, and are able to operate with a limited amount of resources. A real-time operating system is an operating system that guarantees to process events or data by a specific moment in time. A real-time operating system may be single- or multi-tasking, but when multitasking, it uses specialized scheduling algorithms so that a deterministic nature of behaviour is achieved. Such an event-driven system switches between tasks based on their priorities or external events, whereas time-sharing operating systems switch tasks based on clock interrupts.

A computer network is a group of computers that use a set of common communication protocols over digital interconnections for the purpose of sharing resources located on or provided by the network nodes. The interconnections between nodes are formed from a broad spectrum of telecommunication network technologies, based on physically wired, optical, and wireless radio-frequency methods that may be arranged in a variety of network topologies. A Local Area Network (LAN) is a network that connects computers and devices in a limited geographical area, such as a home, school, office building, or closely positioned group of buildings. Each computer or device on the network is a node. Wired LANs are most likely based on Ethernet technology. A Wide Area Network (WAN) is a computer network that covers a large geographic area, such as a city, country,

or spans even intercontinental distances. A WAN uses a communications channel that combines many types of media, such as telephone lines, cables, and air waves.

In this unit, you will study about the Operating System (OS) and NetWare: DOS, UNIX, Windows, and computer networking LAN and WAN.

NOTES

4.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the significance of an Operating System (OS)
- Discuss why an OS is required
- Explain the different types of OS and their commands, such as DOS, UNIX, and Windows
- Elaborate on the networking system LAN, MAN and WAN
- Define what NetWare is

4.2 OPERATING SYSTEM

An Operating System (OS) is the main control program for handling all other programs in a computer. The other programs, usually known as 'application programs', use the services provided by the OS through a well-defined Application Program Interface (API). Every computer necessarily requires some type of operating system that instructs the computer about operations and use other programs installed in the computer. The role of an OS in a computer is similar to the role of the manager in an office for the overall management of the college.

Any computer system can be broadly classified in terms of four component dimensions:

(i) Hardware
(ii) Operating System
(iii) Application Programs (like MS Word, Games, Calculator)
(iv) Users (People who work on the computer)

Figure 4.1 displays the various components of the computer system.

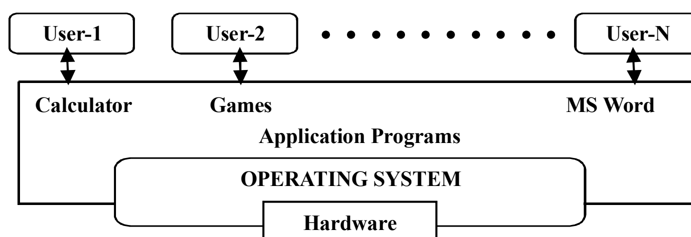


Fig. 4.1 Components of a Computer System

NOTES

Components of an Operating System

An operating system has three primary components:

1. **Control Programs:** As the name implies, these control and maintain the operations of a computer. There Initial Program Loader (IPL) is located in the form of firmware and is stored in the Read Only Memory (ROM) section of the Memory Unit. When the computer is switched on, the electronic chip is automatically activated and the IPL reads the Boot Sector Routine (BSR), which resides in the primary memory of the computer.
2. **System Service Programs**: These support the control programs. Each of these is divided into three portions: Linkage Editor, Librarian and Input/ Output Control System.

   A linkage editor is an editor program that establishes a single module from many modules by resolving cross-references among the modules.
3. **Utility Programs**: These programs run on the operating system in order to carry out various user related commands to manage the software linking the user and the operating system. In MS-DOS, for example, the utility programs are FDISK, FORMAT, ATTRIB, BACKUP, FIND, and others.

Booting Process: In computing, a bootstrapping process called booting (booting up) starts the OS when the computer system is switched on. The first set of operations performed by the computer when switched on is called boot sequence. The main operating system for the computer is loaded by the bootloader.

One can boot an operating system in two conditions: (i) Where there is a single OS installed and (ii) Where there are multiple OSs installed on the computer.

Common terminology

Multi-Tasking: A type of OS that permits multiple programs to be run simultaneously by the same computer. A user of the computer can simultaneously play games while a Word document is being printed (The user is simultaneously working with two different applications–Word and Games). Operating systems supporting multitasking include UNIX and the Windows range.

Multi-Threading: A form of multitasking that permits multiple parts of a software program to be run simultaneously; for example, a user can perform a spell check on a Word document and simultaneously print another Word document. (User is working with two different components (Spell Check) and printing of the same application (Word). Operating systems supporting multitasking include UNIX and Windows.

Multi-Processing: This concerns the use of multiple processors (more than one CPU) to simultaneously execute multiple programs. The inclusion of multiple CPUs in a single computer system improves the performance to a large extent. Multiprocessing involves simultaneous processing by a computer system having

multiple CPUs, whereas multitasking involves simultaneous processing by a computer system with a single CPU. Operating systems supporting multiprocessing include UNIX and Windows NT.

NOTES

Single user: This category of OS does not permit multiple users to use the computer and run programs at the same time. This assumes that at any given time only one user uses the system and runs only one program, i.e., it does not allow two users to concurrently work on the same program. MS DOS is an example.

Multi-user: This class of OS permits multiple users to use the computer and run programs at the same time, e.g., UNIX, Linux, Windows NT.

Functions of an Operating System

The following are the primary functions of an OS:

(i) ***Resource Management:*** Computer resources include main memory (RAM), storage devices (Floppy Disk and Hard Disk Drives), and input and output devices (Keyboard, Mouse, Monitor, Printer). The OS is responsible for:

- Allocating and deallocating memory space as needed by various application programs.
- Functioning as a secondary storage management. All application programs, compilers and loaders are stored in the secondary storage (hard disk).
- Dealing with input and output to and from other connected hardware devices such as printers, hard disks, modems and scanners.

(ii) ***File Management:*** The OS is responsible for the creation and deletion of files/directories and the mapping of these files/directories on to the secondary storage.

(iii) ***Security Management:*** The OS is responsible for protecting the resources and information of a computer system from destruction and misuse.

(iv) ***Operating System Services:*** The OS is responsible for providing a set of services to programs and users of those programs. The main services include:

- Program Execution: It loads the program requested by the user into the memory.
- Error Detection: It generates messages to each application or user about the status of the operations that have been performed. It constantly detects and corrects errors generated by the system.
- Resource Utilization: It ensures efficient utilization of the computer's resources.

NOTES

Check Your Progress

1. What do you mean by an operating system?
2. What are the four classification of computer?
3. Define booting process.
4. What are the functions of operating system?

4.3 TYPES OF OPERATING SYSTEMS

Popular OS include the following:

4.3.1 UNIX

In 1960, the Massachusetts Institute of Technology, AT&T Bell Laboratories and General Electric worked on an operating system, MULTiplexed Information and Computing Service (MULTICS). The MULTICS operating system was designed to run the GE-645 mainframe computers. The operating system was not very successful. Ken Thompson, who was a developer in the AT&T Bell Laboratories, developed a game called Space Travel for the GE-645 mainframe computers. The system was running at a very slow speed. Then, he wrote the game in the assembly language with the help of Dennis Ritchie.

Ken Thompson and Dennis Ritchie included Rudd Canady and formed a team of developers. The aim of the team was to develop an operating system with the support for file system and multitasking for the PDP-7 machines. The operating system included a command line interpreter and different utility programs.

In 1970, an operating system named UNiplexed Information and Computing System (UNICS) was developed, which allowed two users to work simultaneously. The name was changed to UNIX from UNICS. UNIX, as an operating system, was developed in the early 1970s at AT&T Bell Laboratories.

Versions of UNIX

In 1970, UNIX, which was developed to work on the PDP-7 machines, was used on the PDP-11/20 machine for adding text formatting programs. A text formatting program, called ***roff*** and a text editor was added to the operating system. The *roff* was the first electronic publishing program with the capability of typesetting. The first manual of UNIX was printed on 3 November, 1973. This was the first version of UNIX.

In 1973, it was decided to write UNIX in the C programming language. UNIX was made available for universities and commercial firms. The versions of UNIX were determined by the printing of manuals. Some extra features such as increased development speed, and pipes were added in the fourth, fifth and sixth versions of UNIX.

NOTES

The UNIX/32V was released for the Virtual Address eXtension (VAX) systems, which are the 32-bit computing architectural systems in 1978. The version 7, UNIX was released in 1979.

In 1982, AT&T released UNIX System III based on the version 7 UNIX. Similarly, UNIX system V was introduced that contained the VI Editor. AT&T added various features such as file locking, system administration and remote file system to the UNIX system V.

During 1993, the UNIX system V Release 4 was introduced and it was sold to Novell. Novell developed its own version, called UNIXWare. Novell sold all the certification rights of UNIXWare to the X/Open Consortium. In 1996, X/Open merged with Open Software Foundation (OSF) and formed the Open Group.

In 2000, Novell sold the base code of UNIX to Santa Cruz Operation (SCO), which in turn sold it to the Caldera systems, also known as the SCO group. The Sun Microsystems released the OpenSolaris project, which was an open source project containing the system code based on the UNIX system V Release 4.

UNIX is available in various versions. Table 4.1 lists the versions of the UNIX operating system.

Table 4.1 *Versions of UNIX*

Product	***Company***
Xenix	Microsoft Corporation
FreeBSD, NetBSD and OpenBSD	Berkeley Software Design
Solaris	Sun Microsystems
AIX	IBM
HP-UX	Hewlett-Packed/Packard
Digital UNIX	Digital Equipment Corporation
Tru64 UNIX	Digital Equipment Corporation
IRIX	Silicon Graphics
SCO Open Server	Santa Cruz Operation
SCO UNIXWare	Santa Cruz Operation
Linux	Red Hat, SuSE, Caldera and many other companies

The Xenix and the SCO Open Server products are discontinued. The SCO Open Server product is replaced by the SCO UNIXWare product. The FreeBSD, NetBSD, and OpenBSD are commercial products started by the people who created BSD UNIX. The Linux is the clone of UNIX, gaining enormous popularity. The UNIX version that we are using is the FreeBSD product of Berkeley Software Design Company.

NOTES

Introduction to UNIX OS

UNIX is a multi-user and multitasking operating system. In a multi-user environment, the computer can receive the commands from a number of end-users to run programs, access files, and print documents simultaneously. Figure 4.2 shows a multi-user environment.

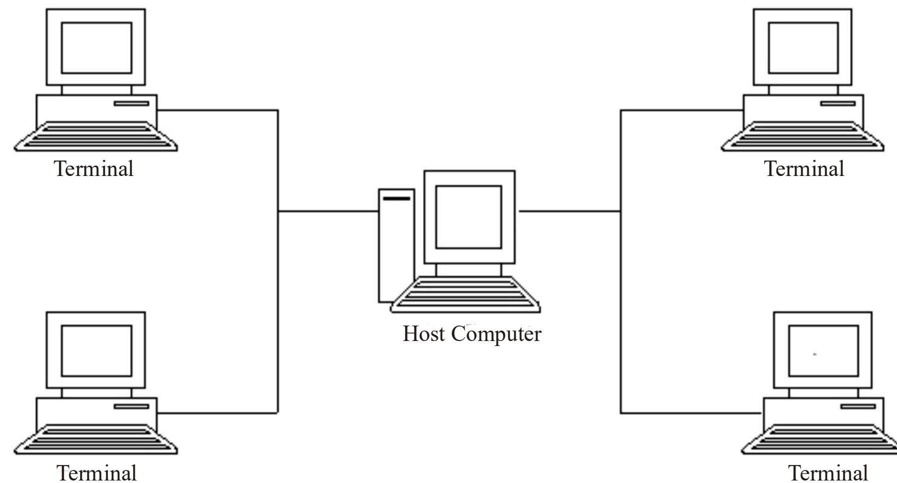


Fig. 4.2 *Multi-User Environment*

The host computer, which has a UNIX Operating system, provides services to the terminals, such as file access services. Four terminals are connected to one host computer and all the terminals are sharing resources from the host computer.

Multitasking feature performs scheduling of work while one task is waiting for input and another task reads input from the hard disk. As a result, you can perform more that one task simultaneously.

Architecture of UNIX

An operating system includes a user interface used for interacting with user and the command interpreter that helps in converting the user language in a high-level language understood by the computer. In UNIX, the command interpreter is not integrated with operating system but as a separate program. The UNIX architecture is divided into different layers such as:

- **Hardware**: Provides the basic hardware interface modules such as memory management, process scheduling, interrupt handling, low-level device control, and process synchronization. These modules help in managing and scheduling the processes received by the user and send these processes to the kernel, the next layer of the UNIX architecture.
- **Kernel**: Controls the resources of the system and forms an interface between shell and hardware. You can interact with the kernel through system calls. System call is a command given by the end user. There is only one kernel in a system. Programs, such as shell and VI Editor interact with the kernel by

NOTES

invoking a well-defined set of system calls. The system calls enables the kernel to perform various operations for the calling programs and it also exchanges the data between the kernel and the program. System calls is also used for file manipulation, process control, information manipulation, and device manipulation.

- **Shell**: Receives a command as an input and sends the command to the kernel for processing. Shell is a program that enables the end user to request the kernel to do the processing of a given command. There can be multiple shells in action in a system, one for each end user who is logged in.
- **Other Program Utilities**: Includes various end user applications such as desktop application, network application, and database application.

Figure 4.3 shows the UNIX system architecture.

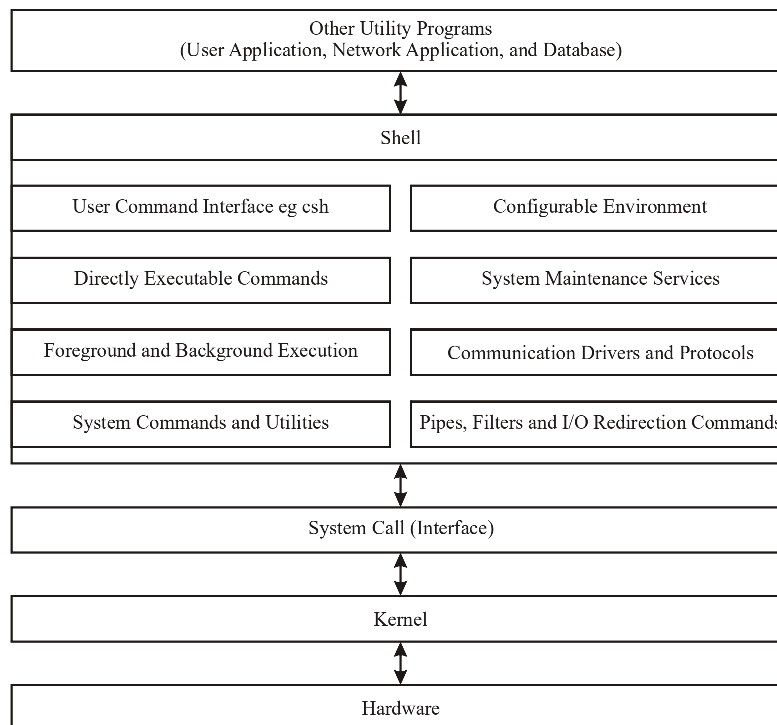


Fig. 4.3 *The UNIX System Architecture*

Features of UNIX

UNIX as an operating system supports all features that are expected in an operating system. UNIX also provides additional features that are not supported in any other operating system. The general and additional features of UNIX operating system are:

- **File and Process**: File and process are two entities that are supported by UNIX. A file contains information such as text, code, or directory structure that you need to save in the computer. The file is stored in the hard disk of the computer at a particular location, which can be easily remembered,

NOTES

whereas a process is the name given to a file or a program that is currently running. UNIX provides various tools that enable you to control a process, change the sequence of the process, and kill the process.

- **Multi-User System**: UNIX supports multitasking system as the kernel is designed to handle multiple processes. A single user can run multiple processes simultaneously; for example, an end user can print a file and edit another file simultaneously. The kernel handles the multiple processes as foreground and background processes. The current process runs in the foreground and the other processes run in the background. The foreground and background processes can be switched according to the requirements. This multitasking feature is an advantage for the programmers, as they do not have to close the editor and run the program; this can be done simultaneously.
- **UNIX Toolkit**: The UNIX toolkit provides various tools that enable you to perform different tasks in UNIX as kernel alone cannot perform every task. The tools that are included in the UNIX toolkit are:
  - o General purpose tools such as VI Editor
  - o Text manipulation utilises filters that are used to retrieve the output from two or more commands simultaneously
  - o Compiler and interpreter
  - o Network administration and system tools such as mailx and pine
- **Pattern Matching**: UNIX supports pattern matching feature that enables you to retrieve the output according to the required pattern. Pattern matching in UNIX can be implemented using a special characters such as * known as metacharacter. The * character denotes multiple characters; for example, if you need to retrieve all the files the names of which start with A, you can give the argument for the list command as A*, instead of specifying all the names of the files.
- **Programming Facility**: UNIX provides a programming facility known as shell that is developed specifically for programmers and not for the users. The shell programming includes all programming features such as variables, loops, and control structures that help you to create a shell script.

Logging on with UNIX

The UNIX operating system provides security for the UNIX users. As multiple users can use the UNIX computer, each user is provided with his/her own login name and password to work in the UNIX computer. The system administrator is the person that is responsible for creating a login name and password for the end users. To enter the login name and password, UNIX provides a login prompt, which is:

```
FreeBSD/i386 (bsd.belmaksit.com) (ttyp0)
login:
```

NOTES

The above login prompt is preceded with the UNIX version installed. You need to enter the login name in front of the login: prompt. After entering the login name press **Enter** from the keyboard, which prompts you to enter the password that was created by the system administrator. Password is the secret code that should not be known to others; for example, if you need to login with the login name, Steve, enter the string at the prompt and press Enter button. Pressing the **Enter** button prompts you to enter the password. The password string will not be visible on the screen. The login screen to login with username Save is:

```
FreeBSD/i386 (bsd.belmaksit.com) (ttyp0)
login: Save
password:
```

Enter the password string and press **Enter** button to login. If the login and password strings are entered correctly then the UNIX provides you with the welcome screen along with the last login time information. The welcome screen is:

```
Last login: Sun Oct 19 11:30:59 from 192.168.0.123
Copyright (c) 1980, 1983, 1986, 1988, 1990, 1991, 1993,
1994
The Regents of the University of California.
All rights reserved.
FreeBSD 4.8-RELEASE (GENERIC) #0: Thu Apr 3
10:53:38 GMT 2003
Welcome to FreeBSD!
$
```

After logging on with UNIX, the system prompts you with the dollar ($) prompt with the blinking cursor.

Working with UNIX Commands

The UNIX system is based on commands, which means that each action in UNIX is performed using commands. UNIX uses Command Line Interface (CLI) to run the commands. The CLI interface is not supported by Microsoft Windows operating system as each task in Windows is performed while clicking a button. The UNIX commands are designed to ensure that maximum amount of output is retrieved from the typed command.

Command Structure

Before using the commands, you must be aware of the command structure that implies how to use the commands. The UNIX command structure is divided into two parts: the command name and the argument used along with the command. The command and argument are separated using a space or a tab to enable the command interpreter to interpret these commands and arguments as a word. The arguments used along with the commands are categorized as:

- Options

- Expressions
- Filenames
- Instructions

NOTES

The syntax for using any command is:

```
Command [options|expression|filenames|instructions]
```

The options argument is the special argument as the options used with a particular command are predefined. The option argument uses a minus sign (–) in front of the option to distinguish it from the filename argument. An example of the use of the option argument is:

```
ls -l
```

where `ls` is the name of the command and `-l` is the option used with the `ls` command. The '–' sign is used with the option argument. Multiple options in a command are used with a single minus sign. An example of the use of multiple minus sign is:

```
ls -l -a -t
```

where `-l`, `-a`, and `-t` are options used with the `ls` command. This command can also be used as:

```
ls -lat
```

The filename argument is used with a command to retrieve the input from a particular file. The filename argument must be at the end of the command after specifying all the options. An example of the use of the filename argument is:

```
ls -lat sample
```

where `-lat` are the options of the `ls` command and sample is the name of the file. The expressions argument is not predefined; you can use any word as an expression along with command. An example of the use of the expression argument is:

```
cat hello
```

where `cat` is a command that is used to display the arguments specified in front of it and `hello` is the expression used with the `cat` command.

UNIX provides certain commands that do not require arguments. For example, the `pwd` command that prints the name of the current working directory does not accept any argument.

UNIX also provides certain commands that may or may not require arguments, which means that providing arguments with these commands is optional. For example, `ls` is a command that is used with and without arguments, according to the requirement.

NOTES

In UNIX certain characters such as pipe sign (|), <, and > operators when used with the commands are not treated as arguments. These characters have some special meaning in UNIX. For example, pipe sign is used when you need to run multiple commands simultaneously.

4.3.2 MS-DOS

Microsoft Disk Operating System (MS DOS) is a single-user operating system built by Microsoft. It was the most commonly used operating system for PC in the 1980s and Microsoft's first commercialized operating system. It was the same operating system that Microsoft developed for IBM's personal computer as a Personal Computer Disk Operating System (PC DOS) and was based on the Intel 8086 family of microprocessors. MS DOS uses Command Line Interface (CLI) that requires knowledge of a large number of commands. Now Graphical User Interface (GUI) based operating systems are becoming popular, hence MS DOS lost its appeal quickly though it was the underlying basic operating system on which early versions of GUI based Windows operating system ran. Even today you will find that Windows Operating Systems continue to use and support MS DOS within a Windows environment. MS DOS was initially released in 1981 and till now eight versions of it have been released. Today, Microsoft has stopped paying much attention to it and is focusing primarily on the GUI-Based Windows Operating Systems.

What is DOS?

Microsoft Disk Operating System (MS DOS) is a single user, single tasking operating system. MS DOS has a command line, text-based/non-graphical user interface commonly referred to as Character-Based User Interface (CUI). When the computer is switched on, a small program checks all internal devices, electronic memory and peripherals. Once this process is completed, MS DOS is loaded.

DOS Prompt

The DOS prompt known as the command prompt looks like C:\> or D:\> where 'C','D' represent the hard drives of the computer system. All commands are typed at the DOS prompt. Enter key is pressed to view the output of the typed command. If the command is correctly typed desired output would be displayed. Otherwise an error message (Bad command or filename/Invalid parameter) is displayed on the screen.

NOTES

DOS Commands: Internal and External

Table 4.2 *Internal DOS Commands*

Command	*Syntax*	*Explanation*	*Example*	*Notes*
DATE	DATE	Displays the system's current date and prompts to enter the new date.	C:\>DATE	The current date is: Fri 05/09/2003 Enter the new date: <mm-dd-yy>:
TIME	TIME	Displays the current time and prompts the user to enter the new time.	C:\>TIME	The current time is: 12:55:25.34 Enter the new time:
VER	VER	Displays the windows version.	C:\>VER	Displays the Windows version installed on your computer.
PROMPT	PROMPT [Text]	Changes the MS DOS command prompt to the specified text. If the command is typed without any parameters then the default prompt setting is restored.	D:\>PROMPT	Changes the prompt to the default setting.
COPY	COPY <Source> <Destination>	Creates a copy of the specified file and places it in the specified location, file will exist at the specified location as well as the source location.	C:\DATA> MOVE HELLO.TXT \LETTER	Creates a copy of HELLO.TXT in the LETTER folder of the C: drive.
REN	REN <Path> <01dfile> <Newfile>	Renames the old file name by the specified new file name.	C:\DATA>REN HELLO.TXT Hi.TXT	Renames 'HELLO.TXT' as 'Hi.TXT'
DFL /FRASF	DEL <Path><Filename>	Deletes the specified file present in the Specified path/location from the hard disk.	C:\DATA>DEL Hi.TXT	Deletes the file 'Hi.TXT' located in the 'DATA' folder of the C: drive.
TYPE	TYPE <Filename>	Displays the contents of a text file.	C:\DATA>TYPE TMP.TXT	Displays the contents of TMP.TXT.
DIR	DHKDrive/ Directory - Name Name>	Displays all the sub-directories and files of the specified drive/directory. It also shows the size of the files and the date and time they were last modified.	C:\>DIRD:	Displays all the contents (files and directories) of the D: drive.
DIR/P	DHKDrive/ Directory -Name>/P	Displays the contents of directory one screen at a time and pauses until any other key is pressed to continue the display.	C:\>DIR DATA/P	Displays the contents of the 'DATA' directory by pausing the screen.
DIR/W	DIR <Drive/ Directory>/W	Displays the contents of the directory width-wise. It omits file size, date and time of creation of file so that more files can be displayed at one time on the screen.	C:\>DIR DATA/W	Displays the contents of the 'DATA' directory width-wise.
DIR/W/P	DIR <Drive/ Directory>/W/P	The Wide and Pause display option can be combined.	C:\>DIR DATA/W/P	Displays the contents of the 'DATA' directory width-wise and by pausing the screen.
CD	CD<Directory- Name> CD\ - Directly takes to the root directory.	Displays the name of the current directory if no parameter is specified with the command. Changes the current directory to the specified directory.	C:\>CDDATA\ SUBDATA	Changes the current directory to 'DATA\SUBDATA'.
MD	MD <Drive/ Directory-Name>	Creates a new directory in the specified location.	C:\>MD 'HELLO'	Creates a directory named 'HELLO' in the C: Drive.
RD	RD <Directory- Name>	Removes the specified directory.	C:>RD HELLO	To remove a directory first you should come to one level above the current directory and then remove command should be given. This command will delete the 'HELLO' directory from the C:drive.
DELTREE	DELTREE <Directory-Name>	Deletes a directory and all the sub-directories and files in it.	C:\>DELTREE TEMP	Prompts the user for confirmation. If user selects 'Y' (Yes) then the directory 'TEMP' and all its sub-directories will be deleted.

Wild Cards

NOTES

Wild card characters can be used in specifying filenames in DOS. There are two types of wild cards. These are: (?,*).

? It is used to represent any single character in the file name.

SYNTAX: C:\ DIR BA?.TXT

Displays all the text files in the C: drive starting with 'BA' and ending with any single character.

Examples: BAT.TXT, BAG.TXT, BAR.TXT, BAD.TXT, etc.

* It is used to represent one or more characters in a file name.

SYNTAX: C:\ DIR CON*.TXT

Displays all the text files in the C: drive starting with 'CON'.

Examples: CONCEPT.TXT, CONCATENATE.TXT, CONTEMPT.TXT, CONSOLE.TXT, etc.

Table 4.3 *External DOS Commands*

Command	*Syntax*	*Explanation*	*Example*	*Notes*
LABEL	LABEL	Makes, changes, or deletes the label of volume of a disk.	C:\>LABEL	Displays the current volume label and volume serial number. Also prompts to enter a new label.
EDIT	EDIT	Starts MS DOS editor, which produces and changes ASCII files.	C:\>EDIT	Opens the MS DOS editor.
ATTRIB	ATTRIB[+A | –A] [+R | –R] [+H | –H] [+S | –S] <file name> + sets an attribute + Clears an attribute A-Archive attribute R-Read only attribute. H-Hidden file attribute S-System file attribute	Displays or changes file attributes.	C:\>ATTRIB+H +R FIRST.TXT	Sets the attributes of 'FIRST.TXT' as Read only and hidden.
XCOPY	XCOPY <Source> <Destination>	Copies files and subdirectories to the specified location.	C:\>XCOPY C:\DATA C:\INFO	Copies the entire contents of the 'DATA' folder to 'INFO' folder. If the 'DATA' folder contains 'ami' subdirectories, then they will also be copied to the 'INFO' folder.
TREE	TREE [Drive:][Path] [/F][/A]	Displays directory paths and files in each subdirectory. /F – Displays file names in each listed directory. /A – Specifies the alternative characters (plus signs, hyphens, etc.) used to draw.	TREE C: TREE D:	Lists a tree listing of the C drive. List a tree listing of the D drive.
DELTREE	DELTREE <Directory-Name>	Deletes a directory and all the sub-directories and files in it.	C:\>DELTREE TEMP	Prompts the user for confirmation. If user selects 'Y' (Yes) then the directory 'TEMP' and all its sub-directories will be deleted.

Concept of File

A file is created with the help of another application program. A user, for instance, may create a text document in Notepad. So the application program in this case in Notepad and the file is a Notepad file.

NOTES

It is a portion of a software program that is used to store data, information, settings, and/or commands used with that program. Examples of files are Word document files, Excel files, PowerPoint presentation files, database files, and so on.

Apart from the normal application program files such as MS Word, MS Excel, MS PowerPoint, a computer system includes other file types also. These are:

- **Batch File:** A batch file allows the users of MS Windows and MS DOS to generate a list of commands for running in sequence after the execution of the batch file. A batch file can be used for running the commands that are frequently run, deleting files, moving files, etc. No special programming skills are required for a simple batch file and can be prepared by users who have a basic knowledge of MS DOS commands.

  An instance of a batch file is a file that is somewhat similar to an icon on the Mac OS or a shortcut in Windows. Similar to a shortcut, batch files can be used for running one or more commands and/or programs through the command line. Another popular instance of a batch file is a simple boot file called autoexec.bat, which is loaded every time the OS is loaded. It works on computers with early Windows. This batch file had all the required commands and programs used for running MS DOS and Windows every time the computer started.
- **Executable File:** It is a file that performs different functions or operations on a computer. Since an executable file is compiled, unlike a data file, it is generally not readable. On an IBM compatible computer, the common executable files are .COM, .BAT, .BIN and .EXE. Other types of executable files can also be there depending on the operating system and its setup.
- **System File:** A file that is being used by an operating system and cannot be deleted or changed without the stopping of the operating system is known as a system file. Since these files are in use by the operating system, generally they cannot be deleted.

A system file is also an attribute that can be added to any file in Microsoft operating systems that allows the OS to know the file is an important system file. Files that are marked as system files will also be hidden files.

Limitations of MS DOS

- It has a text based user interface where the commands have to be typed for each operation that the user wants to perform. The user is expected to remember the commands as well as their syntax.
- It is a single user, single task operating system and the working is limited to one megabyte of memory. 640 KiloBytes of the memory is used for the application program.

- It does not allow using long file names. The user is restricted to eight-character file names with three-character extensions.

4.3.3 Mac OS

NOTES

Mac OS is the operating system designed for the Apple range of personal computers, the Macintosh. Certain important facts about the MAC OS are as follows:

- It was first released in 1984 with the original Macintosh computer and was the first OS to incorporate GUI. In fact, in contrast to the other operating systems available at the time which used a Command Line Interface (CLI), Mac OS was a pure GUI as it had no CLI at all. The philosophy behind this approach to operating system design was to make a system that was user friendly and intuitive where MS DOS and UNIX appeared complicated and challenging to use in comparison.
- Mac OS was originally very hardware-specific, only running on Apple computers using Motorola 68,000 processors. When Apple started building computers using PowerPC processors and hardware, Mac OS was updated to run on these machines. This was the case since the original Mac OS, until Mac OS version 9 was released in 2000. All these versions of Mac OS were pure GUIs.
- The release of OSX (or Mac OS 10) was a significant change in the development of Apple operating systems. OSX was built on UNIX technology and introduced better memory management and multitasking capabilities in the OS. It also introduced a CLI for the first time. Previous Mac OS versions had problems with multiple applications causing them to crash while running simultaneously.
- OSX was originally developed to only run on PowerPC hardware, but since 2006 it has been able to run on Intel or x86 processors and had the following features:
    - o It was the first GUI with focus on usability and simplicity in an operating system.
    - o The intuitive interface and development of publishing and creative software since the first release of Mac OS has made Macintosh computers a favourite in the design and publishing industries.

4.3.4 IBM OS/2

Operating System 2 or OS/2 was a joint effort by IBM and Microsoft for developing a successor to MS DOS and early versions of Microsoft Windows. After the huge success of Windows 3.1, Microsoft decided to part ways with IBM, which decided to develop the OS/2 operating system itself. Certain important facts about IBM OS/2 are as follows:

NOTES

- Introduced in 1987, this operating system for personal computers was intended to provide an alternative to Microsoft Windows for both enterprise and personal users. Though OS/2 looks like Windows 3.1, it has features that are similar to UNIX, particularly the multi-tasking feature and the ability to support multiple users.
- IBM released OS/2 version 3.0 in 1994 and named it OS/2 WARP in order to highlight its new features as well as to strengthen the brand value which was lost due to IBM and Microsoft's rivalry.
- OS/2 was the preferred operating system of various banks for their Automated Teller Machines (ATM) and railways for their Automated Ticket Vending Machines (ATVM).

4.3.5 Linux

Linux is an operating system that is similar to UNIX and was originally developed by Linus Torvalds, a student at the University of Helsinki. The following are certain facts about Linux:

- As the complete source code for Linux is open and available to everyone, it is referred to as Open Source. The user has the freedom to copy and change the program or distribute it.
- Technically, Linux is strictly an OS Kernel (the kernel is the core of an operating system). The first Linux kernel was released to the public in 1991. It had no networking, ran on limited PC hardware and had little device driver support. Later versions of Linux come with a collection of software including GUI, server programs, networking suites and other utilities to make it a more complete OS.
- Generally, an organization will integrate software with the Linux kernel and release what is called a Linux Distribution. Examples of popular Linux distributions are Red Hat, Mandrake SuSE. These organizations are commercial ventures, selling their distributions and developing software for profit.

Linux is primarily used as an OS for network and Internet servers. Recently, it has gained popularity as a desktop OS for general use since the wider inclusion of GUIs and office suite software in distributions. Linux also provides the following:

- **Multi-Tasking/Multiuser:** Linux allows multiple users to run multiple programs on the same system at the same time.
- **Reliable:** A highly reliable and stable OS, it can run for months, even years, without having to be rebooted.
- **TCP/IP Networking Support:** Linux supports most Internet protocols. TCP/IP is built into the kernel itself. TCP/IP is the communication protocol that binds the Internet.

- **High Level Security:** It has many built-in security features to protect the system from unauthorized access. It stores passwords in encrypted form which cannot be decrypted.

4.3.6 Windows

NOTES

Microsoft Windows, commonly referred to as Windows, is a group of several proprietary graphical operating system families, all of which are developed and marketed by Microsoft. Each family caters to a certain sector of the computing industry. Active Microsoft Windows families include Windows NT and Windows IoT (Internet of Things); these may encompass subfamilies, for example Windows Server or Windows Embedded Compact (Windows CE). Defunct Microsoft Windows families include Windows 9x, Windows Mobile and Windows Phone. Microsoft introduced an operating environment named Windows on November 20, 1985, as a graphical operating system shell for MS-DOS in response to the growing interest in Graphical User Interfaces (GUIs). Microsoft Windows came to dominate the world's Personal Computer (PC) market with over 90% market share, overtaking Mac OS, which had been introduced in 1984. Apple came to see Windows as an unfair encroachment on their innovation in GUI development as implemented on products such as the Lisa and Macintosh (eventually settled in court in Microsoft's favour in 1993). On PCs, Windows is still the most popular operating system. However, in 2014, Microsoft admitted losing the majority of the overall operating system market to Android, because of the massive growth in sales of Android smartphones. In 2014, the number of Windows devices sold was less than 25% that of Android devices sold. This comparison, however, may not be fully relevant, as the two operating systems traditionally target different platforms. Still, numbers for server use of Windows (that are comparable to competitors) show one third market share, similar to that for end user use.

Windows 3.X

The first version of Windows 3.0 was released by Microsoft in 1990. Certain important facts about Windows 3.X are as follows:

- It was a graphical interface-based package and not a complete operating system, because it required DOS to be installed first on the computer and only after that could it be loaded and used.
- With the launch of Windows 3.11, huge improvements in terms of usability and performance were seen because the user did not have to remember complex DOS commands, work on a single application at a time or suffer from the limited use of input devices such as a mouse or trackball.
- Some of the prominent features of Windows 3.0 and 3.11 are a supported GUI where programs could be executed just by double clicking on them and most of the system settings could be modified from one point called the Control Panel; it could perform most of the DOS housekeeping commands, such as creating, renaming and deleting directories, copying, moving,

NOTES

renaming, deleting files, formatting disks, etc.; multiple programs in different windows; interchange of data within different applications using a utility called clipboard; support for more options such as fax, drawings, graphical internet browsing. Mixed text and graphical documents, etc.

- Most of the DOS applications could be executed from within the Windows environment and graphical interface was extended to those applications which were designed according to Windows.

Windows 95

Windows 95 was a graphical user interface released by Microsoft Corporation in 1995. It had significant improvements over the earlier version of an operating system distributed by Microsoft under the name of Windows 3.11. Certain features of Windows 95 are as follows:

- In addition to the complete change in the user interface, there were a number of important internal modifications made to the core of the operating system. Windows 95, also known as Windows version 4.0 during its development phase, was one of the most successful operating systems of that time.
- Windows 95 operated independently of MS-DOS rather than in conjunction with it and reduced the use of MS-DOS to only a boot loader for Windows 95. Internet Explorer 4.0 was included in the OEM service release 1 along with TCP/IP support for connecting to the Internet. The final service pack for Windows 95 also included Internet Explorer 5.5.
- Facilitates easy installation of hardware peripherals and software applications through plug 'n' play capabilities under which most devices did not require drivers for using them or can be plugged in without rebooting the system.
- Hybrid compatibility: In terms of support to previous versions of applications and drivers based upon the 16-bit Windows 3.11 file system. Windows 95 was able to support 16-bit programs and 100 per cent native support to DOS programs while managing 32-bit applications simultaneously.

Windows 98

Microsoft released the next version of Windows in 1998. Like its predecessor, Windows 98 supported a hybrid 16/32 bit file access system and better graphical user interface. Certain important facts about Windows 98 are as follows:

- It is often referred to as an operating system that 'Works Better, Plays Better'. Code named 'Memphis' during its development stage, Windows 98 integrated Internet Explorer into the user's desktop to allow its users get a global view of technologies over the World Wide Web (WWW) and enable easy access to it.
- Multiple displays supported using several Visual Display Units (VDUs) simultaneously to increase the capacity of the desktop and support running

NOTES

of different programs on separate monitors. Help on an extensive and easy to use self-help system was provided in its interface.

- In Windows 98, new interfaces such as Universal Serial Bus (USB), Digital Versatile Disk (DVD) and Advanced Configuration and Power Interface (ACPI) were also supported.

Windows 98 SE

The Windows 98 SE (Second Edition) is an improved and enhanced version of Windows 98. Certain important facts about Windows 98 SE are as follows:

- It includes new versions of Microsoft applications as compared to Windows 98, to improve user experience and stability of the operating system.
- Some of the new or improved elements of this operating system are: inclusion of Internet connection sharing, Windows Driver Model (WDM) for Modems, Wake on LAN, Internet Explorer 5.0, integrated support for DVD-ROM drivers, bug free Windows, Microsoft Plus!, support for Web TV and updates for other Microsoft programs, such as NetMeeting, MSN, Microsoft Wallet, Windows media player, etc.

Windows Millennium Edition

Windows ME was released on 14 September, 2000, targeted especially at home PC users. Certain important features of Windows ME are as follows:

- This OS was in continuation to Windows 98 with restricted access to real mode MS DOS shell to improve functionality. Among other changes, Windows ME incorporated was an improved look and feel to the user interface and a system restore option of going back to a previous state of the machine.
- The key features of this operating system were that it had upgraded version of Microsoft products such as Internet Explorer 5.5, Windows Media Player 7, System restore options, applications to easily connect with digital cameras and scanners, Windows Movie, Improved Generic support for USB interface and shell extension of ZIP files into the Windows Explorer.

Windows NT

Microsoft released this version of Windows in 1993. Certain important factors about Windows NT are as follows:

- It increased ease of use and simplified management.
- It used the Windows 95 interface and included advanced network support, trouble-free and better access to the Internet and corporate intranets. With the intent of designing it as an operating system capable of supporting high-level language and at the same time processor independent and support a

NOTES

multiuser and multiprocessing environment, Windows NT had high acceptance in both the home user and professional user markets.

- It became the first Windows OS that combined support for high-end client/server business applications.
- It included new features for performance, security, power of operating system, desktop scalability and dependability.
- It included support for multiprocessor (more than one CPU) architecture.
- Windows NT was geared towards business users and had a rich Application Programming Interface (API), which made it easier to run high-end engineering and scientific applications.
- It also supported full 32-bit system of prosors and memory addressing (technique of transferring data from Memory) unlike the previous versions of Windows which were 16/32 bit hybrids.
- Various versions of Windows NT were released over the years, starting from Windows NT 3.1 in 1993 to Windows NT 4.0 in 1996, after which product development was stopped by Microsoft.

Windows 2000

Microsoft released this version of Windows in 2000. It was an upgrade from Windows NT 4.0 and was designed with the aim of replacing Windows 95, Windows 98 and Windows NT on all business desktops and laptops. Certain factors about Windows 2000 are as follows:

- This version was easy to use, Internet compatible and supported mobile computing. It made hardware installation much easier by including support for a range of new Plug and Play devices, including advanced networking and wireless products, infrared and USB devices.
- The main features of Microsoft Windows 2000 were: Dump capabilities, wherein the operating system gave its users the option of dumping either a part of the memory or the entire contents into a file on the hard drive which helped in saving critical information in case of a system failure; Microsoft Management Control to control the access to administrative tools and system settings, and Recovery Console, to support the distributed file system.

Windows 2003

Windows 2003 was released by Microsoft on 24 April, 2003. Certain important factors about Windows 2003 are as follows:

- This OS was designed and developed over various functional parts of Windows 2000 and Windows XP.
- It boasted better stability, compatibility and security than Windows 2000 and XP. It improved performance of the system by taking advantage of the

NOTES

recent hardware developments, redesigning the system interface and developing better services.

- Major updates of Microsoft in-house applications and services such as Networking, Web Server, Compatibility with Windows NT, etc., were released with this OS.
- It provides support for 64-bit processors, Internet Information Services V6.0, a separate Web Edition of Windows 2003 specially designed as a web server, tighter security measures over previous versions of Windows using built-in firewall; support of a hardware based monitoring system called 'watchdog timer' which could monitor the server for hang-ups and freezes; virtual disk services for offsite storage and support for multiple roles such as that of a Web server, print server and storage server.

Windows XP

Windows XP was first released on 25 October, 2001 and since then over 600 million copies have sold worldwide. It is a successor to both Windows 2000 and Windows ME and the first OS aimed at home users built on the Windows NT kernel and architecture. Certain important facts about Windows XP are as follows:

- Owing to the integration of multiple technologies from various operating systems, it gained wide popularity among home and business desktop, notebooks and media centre users. As acknowledged by most Windows XP users as well as Microsoft Corporation, this version of Windows is the most stable and efficient OS released by Microsoft yet.
- Enhanced support to drivers for hardware devices connected to the computer using an improved version of Windows Image Acquisition (WIA). An option to revert back to driver changes has also been added in the control panel.
- CD-burning capabilities are integrated into the Windows XP OS. This technology is adopted from Roxio and does not require a third party software to be installed in order to burn CDs and DVDs.
- Simultaneous user login into the operating system enables multiuser switching between tasks and does not require one user to close all applications before letting another user log in.
- Remote assistance helps a Windows XP user take control of another Windows XP machine by using a network or the internet. This is very helpful in remotely fixing problems without being physically present around the machine.
- Improvement in Fonts using ClearType technology made it easier and more attractive to work in Windows XP.
- Power Management functions are drastically improved and default standard changed to ACPI (Advance Configuration and Power Interface)

NOTES

from APM (Advanced Power Management). With ACPI, according to battery status processor speeds could be reduced or increased instantly, USB devices could be individually isolated and suspended to save power or screen brightness could be adjusted to increase battery life.

- This was the first OS from Microsoft which required product activation to fight piracy.

Windows Vista

The most recent in the line of Microsoft Windows personal computer operating system, Windows Vista, codenamed Longhorn, was developed to succeed Windows XP. Microsoft started the development of Windows Vista five months after releasing Windows XP and the work continued till November, 2006, when Microsoft announced its completion, ending the longest development cycle of an operating system. Since the original idea of building Longhorn from the code of Windows XP was scrapped, it was built on Windows 2003 SP1, with several developments including all new graphical interface named Windows Aero, refined and faster search capabilities, an array of new tools such as Windows DVD Maker, integrated Windows Media Centre in the Vista Home Premium and Vista Ultimate Editions, print, audio, display subsystems and redesigned networking.

The key features of Windows Vista (all versions) are stated below:

- Increased level of communication by the use of peer-to-peer technology between computers on a home network for simplifying the process of sharing files and digital media between the computers and the attached devices.
- It includes the 3.0 version of the .NET Framework with the aim of making application writing significantly easier for software developers than with Windows API.
- Windows Aero, the new graphical interface of Windows Vista OS, is an aesthetically driven GUI with transparencies, live thumbnails and icons. The overall look and feel of the GUI is pleasing to the eye and convenient to work with.
- Instant Search is a new feature of Windows Vista which is significantly faster and returns better in-depth results for files and folders on desktop.
- Windows Sidebar is a panel where selected Windows gadgets are located. These gadgets update the user on various topics such as stock indexes, sports score, currency exchange rate, etc., and can be customized according to user requirements.
- Windows Internet Explorer 7 incorporates tabbed browsing, Anti-Phishing filtering and works in isolation from other applications using a protected mode.
- The Backup and Restore application provides the user with the ability to backup and restore application at schedule periodic intervals of files and

NOTES

folders present on their computers. Backups are stored on the basis of changes made to the data and incremented automatically to the existing backup. The option to completely backup all data on the PC is also available in selected editions of Windows Vista, wherein an image can be created on hard drives or DVDs. In case of a hardware or software failure, complete PC Restore can be easily performed and data loss can be prevented.

- Windows DVD Maker brings native support to Windows Movie Maker for creating custom DVDs based on user's content. Operations like designing title, menu, soundtrack, video, zoom and pan motion effects on slides or pictures can be easily performed.
- Windows Media Centre, which used to be a separate edition of Windows XP known as Windows XP Media Centre. Edition now comes integrated with Windows Vista in the Home Premium and Ultimate edition.
- Windows Mobility Centre is a panel for controlling that integrates the most appropriate information pertaining to mobile computing (sound, brightness, power scheme selection/ battery level, wireless network, presentation settings, screen orientation, etc.).

Windows CE

The Windows Embedded Compact (CE) is an operating system optimized for devices with minimum hardware resources, such as embedded devices and handhelds. It integrates advance and reliable real-time capabilities with Windows technology. The kernel of this OS is not just a trimmed down version of desktop Windows, but in fact, it is a brand new kernel which can run on less than a megabyte of memory. Besides, the advantage of performing on a minimum specification, it is also an OS which satisfies the prerequisites of a real-time operating system. Another distinct feature of Windows CE is that it was made available in a source code form to several hardware manufacturers so that they could modify the OS to adjust with their hardware and also to the general public. Since Windows CE was developed as a component based and embedded operating system, it has been used as a basis in the development of several mobile operating systems such as AutoPC, PocketPC, Windows Mobile, Smartphone, etc., and also embedded into games consoles such as Microsoft Xbox.

Some of the key features of Windows CE are given below:

- Connectivity with a wide array of options such as wireless communication, infrared, dial-up networking or Ethernet network connections. With advance security encryption and continuous synchronization with personal computers and other handheld devices.
- Availability of business applications such as Excel, Word, Outlook, Power point, etc., on the move is a great advantage towards improving productivity and easier access to information.

NOTES

- The option to connect to different types of printers and support for multiple modes of connection such as, serial port or infrared.
- New support for programming languages with ActiveX, DirectX, Hypertext Markup Language (HTML), Java Virtual Machines(JVM), Visual Basic Script (VBS), etc.
- Easy to use GUI with a cascading Start menu, colour and greyscale screens, customization of command bars, support for international character set, ability to display VGA graphics using an external display, true type fonts, etc.
- Starting from Windows CE 5.0 onwards, the support for Voice over Internet Protocol (VoIP) was also added.
- Besides smooth browsing using the integrated Internet Explorer, other features such as Remote Desktop, Web Services on Devices (WSDAPI), Windows media player updates, etc., are also available in Windows CE.

Check Your Progress

5. List the versions of UNIX OS.
6. Define the wild cards along with its examples.
7. Why is pattern matching feature useful in UNIX?
8. What is meant by DOS Prompt?
9. What are the important facts about Windows 98 SE?

4.4 AN OVERVIEW OF NETWORKING

Earlier, a computer network consisted of mainframes in an enclosure. You could see input in the form of a punch card. These were read by card readers and output was in the form of printed results via local printers. Local terminals were mostly used for control and programming input. All processing was on a batch basis, rather than being interactive. In other words, the input was transmitted in a batch from a card reader over a short circuit to the processor. The processor processed the program in a batch and the output to the printer was in a batch. The first true mainframe was the IBM 360, introduced in 1964. Over time, input to the mainframe was extended to multiple users at dumb terminals that connected to ports on the mainframe through terminal controllers, or cluster controllers.

In parallel to the development of data networking, the computers began to change. Computers became more powerful as processor speeds increased with the development of faster microprocessors on silicon chips. Memory became more available as chip technology and hard drive technology improved. Additionally, computers became smaller and less expensive, to the point that the typical desktop PC is equivalent to an early mainframe that would have filled a moderate-size

office building. As a result, all computing power and storage capability on all these desktops would lead to a need to network with those devices within the workplace. It has been estimated that majority of data transfer is confined to the workplace, while only a small percentage travels to remote places. Hence, it is clear that PC users need to share access to hosts, databases, printers, etc. LANs provide a solution to that requirement.

Robert M. Metcalfe and his associates at the Xerox Palo Alto Research Centre (Xerox PARC) first conceived LAN technology. Later on, Xerox commercialized the technology and named it The Xerox Wire. When Digital Equipment Corporation (DEC), Intel and Xerox corporation decided to standardize the technology in 1979, they eventually named it to Ethernet. Ethernet quickly became a de facto standard. Ethernet and LANs were officially recognized when the IEEE established Project 802 at the request of its members. In the end of 1982, the first standard was published and circulated. Ethernet, clearly, is still the most popular LAN Standard.

Computers are connected by many different technologies. A network is an interconnection between two or more computers in a peer-to-peer or client to server manner or fashion usually over a shared and virtual connection. In other words, networks provide the connection between computer resources in order to accommodate the flow of infor-mation. This is just the opposite of the old terminal-to-host hardwired connection. Although a network can support terminal-to-host connections through terminal emulators or terminal server, it offers a lot more flexibility in switching connections. The disadvantage of this explosion in terms of sharing information arises when one computer wishes to share its information system with another which has different network protocols and different network technology. As a result, even if you could agree on a type of network technology to physically interconnect the two computers at different locations, your applications still would not be able to communicate with each other because of the different protocols.

A very basic question arises about the requirement of networks. This may be justified with the help of following points:

- Sharing of resources can be done easily.
- Reliability — There is no central computer, so if one breaks down you can use others.
- Networks allow us to be mobile.

The term networking applies to:

- The exchange of information among individuals, groups, or institutions.
- The process of electronic voice or data communications.

Communication networks are broadly categorized into three categories:

4.4.1 Local Area Network (LAN)

Local Area Network technology connects people and machines within a site. A Local Area Network (LAN) is a network that is restricted to a relatively small

NOTES

area as shown in Figure 4.4. Local Area Networks (LANs) can be defined as privately owned networks offering reliable high speed communication channels that are optimized to connect information processing equipment in a small and restricted geographical area, namely, an office, a building, a complex of buildings, a school or a campus.

A LAN is a form of local (limited-distance), shared packet network for computer communications. LANs interconnect computers and peripherals over a common medium so that users are able to share access to host computers, databases, files, applications, and peripherals. They can also provide a connection to other networks either through a computer, which is attached to both networks, or through a dedicated device called a gateway.

LANs are classified depending on the topology, access methods, signalling methods, transmission medium and transmission mode.

The control of the network is of two types viz., centralized and distributed. With centralized control, access to the network and allocation of channel is controlled by one node, such as a dedicated communications processor or switch. When control is distributed, nodes have the ability to establish connections and access the network channel independently, according to an accepted set of rules.

Most LANs are based on simple structured topologies, like the ring, bus, or star.

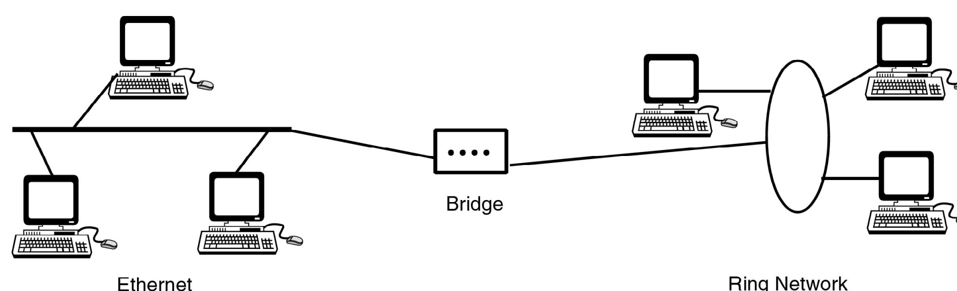


Fig. 4.4 *Local Area Network (LAN)*

The components used by LANs can be categorized into hardware, cabling standards, and protocols. Various LAN protocols are Ethernet, Token Ring: TCP/IP, SMB, NetBIOS and NetBeui, IPX/SPX, Fibre Distributed Data Interchange (FDDI) and Asynchronous Transfer Mode (ATM).

4.4.2 Metropolitan Area Network (MAN)

A Metropolitan Area Network (MAN) covers large geographic areas such as towns, cities or districts. By linking or interconnecting smaller networks within a large geographic area, information is conveniently distributed throughout the network. Local libraries and government agencies often use a MAN to establish a link with private industries and citizens. It may also connect MANs together within a larger area than a LAN. The geographical limit of a MAN may span a city. Figure 4.5 depicts how a MAN may be available within a city.

NOTES

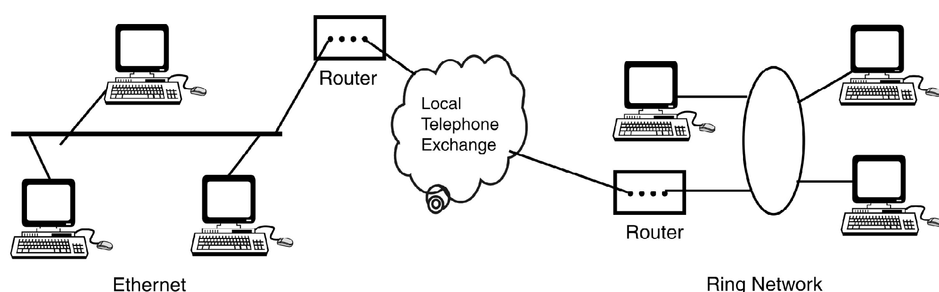


Fig. 4.5 Metropolitan Area Network (MAN)

In MAN, different LANs are connected through a local telephone exchange. Some of the widely used protocols for MAN are RS-232, X.25, Frame Relay, Asynchronous Transfer Mode (ATM), ISDN (Integrated Services Digital Network), OC-3 lines (155 Mbps), ADSL (Asymmetrical Digital Subscriber Line) etc. These protocols are quite different from those used for LANs.

4.4.3 Wide Area Network (WAN)

This technology connects sites that are in diverse locations. Wide Area Networks (WANs) connect larger geographic areas, such as New Delhi, India, or the world. The geographical limit of WAN is unlimited. Dedicated transoceanic cabling or satellite uplinks may be used to connect this type of network. Hence, a WAN may be defined as a data communications network covering a relatively broad geographical area to connect LANs together between different cities with the help of transmission facilities provided by common carriers, such as telephone companies. WAN technologies operate at the lower three layers of the OSI reference model. These are the physical data link and network layers.

Figure 4.6 explains the WAN, which connects many LAN together. It also uses switching technology provided by local exchange and long distance carrier.

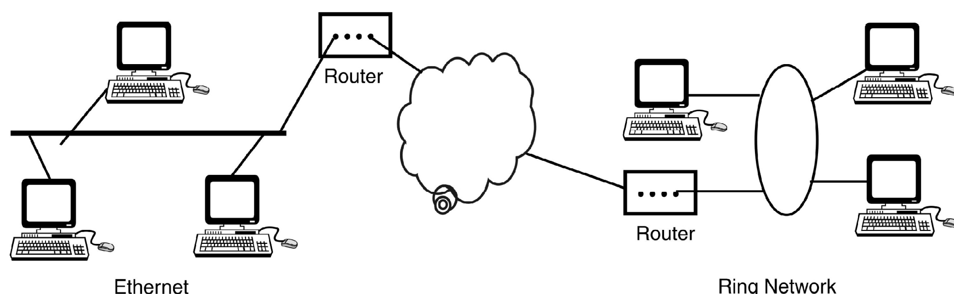


Fig. 4.6 Wide Area Network (WAN)

Packet switching technologies such as Asynchronous Transfer Mode (ATM), Switched Multimegabit Data Service (SMDS), Frame Relay and X.25 are used to implement WAN along with statistical multiplexing to allow devices to use and share these circuits.

The difference between MAN and WAN may be understood only from the services being used by them. WAN uses both the local and long distance carrier

while MAN uses only local carrier. Hardware and protocols are same as in case of MAN.

NOTES

There is a lot of confusion between LAN technology and WAN technology. The answer lies in how data is switched. Switching techniques are described subsequently in this chapter. It is the LAN/(WAN) integration that makes the network work. After all, people and machines not only need to be accessible locally, but from different sites as well.

4.5 NETWARE

NetWare is a discontinued computer network operating system developed by Novell, Inc. It initially used cooperative multitasking to run various services on a personal computer, using the IPX network protocol.

The original NetWare product in 1983 supported clients running both CP/M and MS-DOS, ran over a proprietary star network topology and was based on a Novell-built file server using the Motorola 68000 processor, but the company soon moved away from building its own hardware, and NetWare became hardware-independent, running on any suitable Intel-based IBM PC compatible system, and a wide range of network cards. From the beginning NetWare implemented a number of features inspired by mainframe and minicomputer systems that were not available in its competitors.

In 1991, Novell introduced cheaper peer-to-peer networking products for DOS and Windows, unrelated to their server-centric NetWare. These were NetWare Lite 1.0 (NWL), and later Personal NetWare 1.0 (PNW) in 1993.

In 1993, the main NetWare product line took a dramatic turn when version 4 introduced NetWare Directory Services (NDS), a global directory service similar to the Active Directory that Microsoft would release seven years later. This, along with a new e-mail system (GroupWise), application configuration suite (ZENworks), and security product (BorderManager) were all targeted at the needs of large enterprises.

By 2000, however, Microsoft was taking more of Novell's customer base and Novell increasingly looked to a future based on a Linux kernel. The successor to NetWare, Open Enterprise Server (OES), released in March 2005, offered all the services previously hosted by NetWare 6.5, but on a SUSE Linux Enterprise Server; the NetWare kernel remained an option until OES 11 in late 2011.

NetWare dominated the Network Operating System (NOS) market from the mid-1980s through the mid- to late-1990s due to its extremely high performance relative to other NOS technologies. Most benchmarks during this period demonstrated a 5:1 to 10:1 performance advantage over products from Microsoft, Banyan, and others. One noteworthy benchmark pitted NetWare 3.x

NOTES

running NFS services over TCP/IP (not NetWare's native IPX protocol) against a dedicated Auspex NFS server and an SCO Unix server running NFS service. NetWare NFS outperformed both 'native' NFS systems and claimed a 2:1 performance advantage over SCO UNIX NFS on the same hardware.

History of NetWare

A networking card with a sticker indicating certification with NetWare. NetWare evolved from a very simple concept: file sharing instead of disk sharing. In 1983 when the first versions of NetWare originated, all other competing products were based on the concept of providing shared direct disk access. Novell's alternative approach was validated by IBM in 1984, which helped promote the NetWare product.

Novell NetWare shared disk space in the form of NetWare volumes, comparable to DOS volumes. Clients running DOS would run a special Terminate and Stay Resident (TSR) program that allowed them to map a local drive letter to a NetWare volume. Clients had to log into a server in order to be allowed to map volumes, and access could be restricted according to the login name. Similarly, they could connect to shared printers on the dedicated server, and print as if the printer was connected locally.

i. At the end of the 1990s, with Internet connectivity booming, the Internet's TCP/IP protocol became dominant on LANs. Novell had introduced limited TCP/IP support in NetWare 3.x (circa 1992) and 4.x (circa 1995), consisting mainly of FTP services and UNIX-style LPR/LPD printing (available in NetWare 3.x), and a Novell-developed webserver (in NetWare 4.x). Native TCP/IP support for the client file and print services normally associated with NetWare was introduced in NetWare 5.0 (released in 1998).

During the early to mid-1980s Microsoft introduced their own LAN system in LAN Manager, based on the competing NBF protocol. Early attempts to muscle in on NetWare failed, but this changed with the inclusion of improved networking support in Windows for Workgroups, and then the hugely successful Windows NT and Windows 95. NT, in particular, offered services similar to those offered by NetWare, but on a system that could also be used on a desktop, and connected directly to other Windows desktops where NBF was now almost universal.

Features of NetWare

1. NetWare Server Maintenance

(i) Supported Operating Systems: Novell NetWare 3.x, 4.x, 5.x, 6.x, Novell OES (Linux) 1.x, 2.x.

(ii) Network Monitor.

NOTES

(iii) Server Filtering.

(iv) Extended Connection to Servers and User Authentication.

(v) Connection and Authentication Logs.

(vi) Remote Execution of the Server's Console Commands.

(vii) Blocking and Unblocking the Server.

(viii) Server Shutdown.

2. Backup

(i) Creation of full or partial e-Directory Images (NDS/Bindery Object Database, GroupWise, Object Attributes, Group Settings, User Accounts, Inherited Rights Filters, Trustee Rights, File Structure, etc.).

(ii) Logging of backup operations.

3. Image

(i) Viewing and editing of server images outside the network.

(ii) Browsing history.

(iii) Password protection.

(iv) Image signing and modifications counter.

(v) Importing and exporting images in XML format.

4. Recovery

(i) Complete or partial recovery of e-Directory.

(ii) Transfer of objects and permissions to other NetWare servers of any version.

(iii) Merging of several trees on a single server.

(iv) Cloning servers.

(v) Migration of the Bindery server to the NDS platform and back.

(vi) Logging of recovery operations.

5. Creating Reports

(i) Complete or selective generation of listings for containers, groups, users and the server's file system.

(ii) Creation of aggregate reports for several servers.

(iii) Numerous options for constructing your own reports.

(iv) Creating report profiles.

6. Monitoring Connections

(i) Creation of an instant snapshot of server connections.

(ii) Displaying the list of users connected to the server and connections.

(iii) Displaying the list of file resources occupied on the server.
(iv) Detecting clients' network addresses and their login times.
(v) Detecting active users with the administrator rights.
(vi) Displaying changes.
(vii) Disconnecting separate users and connections from the server.
(viii) Sending short text messages.

7. Quick Navigation

Convenient and thoroughly developed interface allowing you to find the information you need in no time.

Check Your Progress

10. Differentiate between LAN and WAN.
11. Define NetWare.

4.6 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. An Operating System (OS) is the main control program for handling all other programs in a computer. The other programs, usually known as 'application programs', use the services provided by the OS through a well-defined pplication Program Interface (API).
2. Any computer system can be broadly classified in terms of four component dimensions:
   - Hardware
   - Operating system
   - Application programs (like MS Word, Games, Calculator)
   - Users (People who work on the computer)
3. In computing, a bootstrapping process called booting (booting up) starts the OS when the computer system is switched on. The first set of operations performed by the computer when switched on is called boot sequence. The main operating system for the computer is loaded by the boot-loader.
4. The functions of an OS:
   - Resource Management: Computer resources include main memory (RAM), storage devices (floppy disk and hard disk drives), and input and output devices (keyboard, mouse, monitor, printer).

NOTES

- File Management: The OS is responsible for the creation and deletion of files/ directories and the mapping of these files/directories on to the secondary storage.
- Security Management: The OS is responsible for protecting the resources and information of a computer system from destruction and misuse.
- Operating System Services: The OS is responsible for providing a set of services to programs and users of those programs.

5. Versions of UNIX OS are:

Product	***Company***
Xenix	Microsoft Corporation
FreeBSD, NetBSD and OpenBSD	Berkeley Software
Design Solaris	Sun Microsystems
AIX	IBM
HP-UX	Hewlett-Packed/Packard
Digital UNIX	Digital Equipment Corporation
Tru64 UNIX	Digital Equipment Corporation
IRIX	Silicon Graphics
SCO Open Server	Santa Cruz Operation
SCO UNIXWare	Santa Cruz Operation
Linux	Red Hat, SuSE, Caldera and many other Companies

6. Wild card characters can be used in specifying filenames in DOS. There are two types of wild cards. These are: (?,*).

   ? : It is used to represent any single character in the file name.

   SYNTAX: C:\DIR BA?.TXT

   Displays all the text files in the C: drive starting with 'BA' and ending with any single character.

   Examples: BAT.TXT, BAG.TXT, BAR.TXT, BAD.TXT, etc.

   *** :** It is used to represent one or more characters in a file name.

   SYNTAX: C:\DIR CON*.TXT

7. UNIX supports pattern matching feature that enables you to retrieve the output according to the required pattern. Pattern matching in UNIX can be implemented using a special characters such as * known as metacharacter. The * character denotes multiple characters; for example, if you need to retrieve all the files the names of which start with A, you can give the argument for the list command as A*, instead of specifying all the names of the files.

NOTES

8. The DOS prompt known as the command prompt looks like C:\> or D:\> where 'C','D' represent the hard drives of the computer system. All commands are typed at the DOS prompt. Enter key is pressed to view the output of the typed command. If the command is correctly typed desired output would be displayed. Otherwise an error message (Bad command or filename/Invalid parameter) is displayed on the screen.
9. Certain important facts about Windows 98 SE are as follows:
   - It includes new versions of Microsoft applications as compared to Windows 98, to improve user experience and stability of the operating system.
   - Some of the new or improved elements of this operating system are: inclusion of Internet connection sharing, Windows Driver Model (WDM) for Modems, Wake on LAN, Internet Explorer 5.0, integrated support for DVD-ROM drivers, bug free Windows, Microsoft Plus!, support for Web TV and updates for other Microsoft programs, such as NetMeeting, MSN, Microsoft Wallet, Windows Media Player, etc.
10. A LAN is a form of local (limited-distance), shared packet network for computer communications. LANs interconnect computers and peripherals over a common medium so that users are able to share access to host computers, databases, files, applications, and peripherals. They can also provide a connection to other networks either through a computer, which

    ttached to both networks, or through a dedicated device called a gateway. A WAN may be defined as a data communications network covering a relatively road geographical area to connect LANs together between different cities wi h the help of transmission facilities provided by common carriers, such as telephone companies. WAN technologies operate at the lower three lay rs of the OSI reference model. These are the physical data link and n twork layers. It also uses switching technology provided by local exchage and long distance carrier.
11. NetWare is a discontinued computer network operating system developed by Novell, Inc. It initially used cooperative multitasking to run various services on a personal computer, using the IPX network protocol.

4.7 SUMMARY

- An Operating System (OS) is the main control program for handling all other programs in a computer. The other programs, usually known as 'application programs', use the services provided by the OS through a well-defined Application Program Interface (API).
- Any computer system can be broadly classified in terms of four component dimensions:
  (i) Hardware

(ii) Operating system

(iii) Application programs (like MS Word, Games, Calculator)

(iv) Users (People who work on the computer)

NOTES

- Booting Process: In computing, a bootstrapping process called booting (booting up) starts the OS when the computer system is switched on. The first set of operations performed by the computer when switched on is called boot sequence. The main operating system for the computer is loaded by the boot-loader.
- Functions of an Operating System

  The following are the primary functions of an OS:

  (i) Resource Management: Computer resources include main memory (RAM), storage devices (floppy disk and hard disk drives), and input and output devices (keyboard, mouse, monitor, printer).

  (ii) File Management: The OS is responsible for the creation and deletion of files/ directories and the mapping of these files/directories on to the secondary storage.

  (iii) Security Management: The OS is responsible for protecting the resources and information of a computer system from destruction and misuse.

  (iv) Operating System Services: The OS is responsible for providing a set of services to programs and users of those programs.
- In 1970, UNIX, which was developed to work on the PDP-7 machines, was used on the PDP-11/20 machine for adding text formatting programs. A text formatting program, called *roff* and a text editor was added to the operating system. The '*roff*' was the first electronic publishing program with the capability of typesetting. The first manual of UNIX was printed on 3 November, 1973. This was the first version of UNIX.
- UNIX is a multi-user and multitasking operating system. In a multi-user environment, the computer can receive the commands from a number of end-users to run programs, access files, and print documents simultaneously.
- Microsoft Disk Operating System (MS- DOS) is a single-user operating system built by Microsoft. It was the most commonly used operating system for PC in the 1980s and Microsoft's first commercialized operating system. It was the same operating system that Microsoft developed for IBM's personal computer as a Personal Computer Disk Operating System (PC- DOS) and was based on the Intel 8086 family of microprocessors. MS-DOS uses Command Line Interface (CLI) that requires knowledge of a large number of commands.
- The DOS prompt known as the command prompt looks like C:\> or D:\> where 'C', 'D' represent the hard drives of the computer system. All commands

NOTES

are typed at the DOS prompt. Enter key is pressed to view the output of the typed command. If the command is correctly typed desired output would be displayed. Otherwise an error message (Bad command or filename/Invalid parameter) is displayed on the screen.

- Wild card characters can be used in specifying filenames in DOS. There are two types of wild cards. These are: (?,*).
- Mac OS is the operating system designed for the Apple range of personal computers, the Macintosh. It was first released in 1984 with the original Macintosh computer and was the first OS to incorporate GUI. In fact, in contrast to the other operating systems available at the time which used a Command Line Interface (CLI), Mac OS was a pure GUI as it had no CLI at all. The philosophy behind this approach to operating system design was to make a system that was user friendly and intuitive where MS-DOS and UNIX appeared complicated and challenging to use in comparison.
- IBM OS/2: Operating System 2 or OS/2 was a joint effort by IBM and Microsoft for developing a successor to MS DOS and early versions of Microsoft Windows. After the huge success of Windows 3.1, Microsoft decided to part ways with IBM, which decided to develop the OS/2 operating system itself. Introduced in 1987, this operating system for personal computers was intended to provide an alternative to Microsoft Windows for both enterprise and personal users. Though OS/2 looks like Windows 3.1, it has features that are similar to UNIX, particularly the multi-tasking feature and the ability to support multiple users.
- Linux is an operating system that is similar to UNIX and was originally developed by Linus Torvalds, a student at the University of Helsinki.
- Linux is primarily used as an OS for network and Internet servers. Recently, it has gained popularity as a desktop OS for general use since the wider inclusion of GUIs and office suite software in distributions. Linux also provides the following:
    (i) Multi-Tasking/ Multiuser: Linux allows multiple users to run multiple programs on the same system at the same time.
    (ii) Reliable: A highly reliable and stable OS, it can run for months, even years, without having to be rebooted.
    (iii) TCP/IP Networking Support: Linux supports most Internet protocols. TCP/IP is built into the kernel itself. TCP/IP is the communication protocol that binds the Internet.
    (iv) High Level Security: It has many built-in security features to protect the system from unauthorized access. It stores passwords in encrypted form which cannot be decrypted.
- Active Microsoft Windows families include Windows NT and Windows IoT (Internet of Things); these may encompass subfamilies, e.g. Windows

NOTES

Server or Windows Embedded Compact (Windows CE). Defunct Microsoft Windows families include Windows 9x, Windows Mobile and Windows Phone. Microsoft introduced an operating environment named Windows on November 20, 1985, as a graphical operating system shell for MS-DOS in response to the growing interest in Graphical User Interfaces (GUIs). Microsoft Windows came to dominate the world's Personal Computer (PC) market with over 90% market share, overtaking Mac OS, which had been introduced in 1984. Apple came to see Windows as an unfair encroachment on their innovation in GUI development as implemented on products such as the Lisa and Macintosh (eventually settled in court in Microsoft's favor in 1993).

- Different versions of Windows are as follows:
  - (i) The first version of Windows 3.0 was released by Microsoft in 1990.
  - (ii) Windows 95 was a graphical user interface released by Microsoft Corporation in 1995.
  - (iii) The Windows 98 SE (Second Edition) is an improved and enhanced version of Windows 98.
  - (iv) Windows ME was released on 14 September, 2000, targeted especially at home PC users.
  - (v) Windows NT was geared towards business users and had a rich Application Programming Interface (API), which made it easier to run high-end engineering and scientific applications.
  - (vi) Microsoft released this version of Windows in 2000. It was an upgrade from Windows NT 4.0 and was designed with the aim of replacing Windows 95, Windows 98 and Windows NT on all business desktops and laptops.
  - (vii) Windows 2003 was released by Microsoft on 24 April, 2003.
  - (viii) Windows XP was first released on 25 October, 2001 and since then over 600 million copies have sold worldwide. It is a successor to both Windows 2000 and Windows ME and the first OS aimed at home users built on the Windows NT kernel and architecture.
  - (ix) The most recent in the line of Microsoft Windows personal computer operating system, Windows Vista, codenamed Longhorn, was developed to succeed Windows XP. Microsoft started the development of Windows Vista five months after releasing Windows XP and the work continued till November, 2006, when Microsoft announced its completion, ending the longest development cycle of an operating system.
  - (x) The key features of Windows CE are - Connectivity with a wide array of options, such as wireless communication, infrared, dial-up networking

or Ethernet network connections. With advance security encryption and continuous synchronization with personal computers and other handheld devices. Availability of business applications such as Excel, Word, Outlook, Power point, etc., on the move is a great advantage towards improving productivity and easier access to information. The option to connect to different types of printers and support for multiple modes of connection such as, serial port or infrared. New support for programming languages with ActiveX, DirectX, HyperText Markup Language (HTML), Java Virtual Machines(JVM), Visual Basic Script (VBS), etc. Easy to use GUI with a cascading Start menu, colour and greyscale screens, customization of command bars, support for international character set, ability to display VGA graphics using an external display, true type fonts, etc. Starting from Windows CE 5.0 onwards, the support for Voice over Internet Protocol (VoIP) was also added.

- Robert M. Metcalfe and his associates at the Xerox Palo Alto Research Centre (Xerox PARC) first conceived LAN technology. Later on, Xerox commercialized the technology and named it The Xerox Wire. When Digital Equipment Corporation (DEC), Intel and Xerox corporation decided to standardize the technology in 1979, they eventually named it to Ethernet. Ethernet quickly became a de facto standard. Ethernet and LANs were officially recognized when the IEEE established Project 802 at the request of its members. In the end of 1982, the first standard was published and circulated. Ethernet, clearly, is still the most popular LAN Standard.
- A LAN is a form of local (limited-distance), shared packet network for computer communications. LANs interconnect computers and peripherals over a common medium so that users are able to share access to host computers, databases, files, applications, and peripherals. They can also provide a connection to other networks either through a computer, which is attached to both networks, or through a dedicated device called a gateway**.**
- In MAN, different LANs are connected through a local telephone exchange. Some of the widely used protocols for MAN are RS-232, X.25, Frame Relay, Asynchronous Transfer Mode (ATM), ISDN (Integrated Services Digital Network), OC-3 lines (155 Mbps), ADSL (Asymmetrical Digital Subscriber Line), etc. These protocols are quite different from those used for LANs.
- WAN may be defined as a data communications network covering a relatively broad geographical area to connect LANs together between different cities with the help of transmission facilities provided by common carriers, such as telephone companies. WAN technologies operate at the lower three layers of the OSI reference model. These are the physical data link and network layers.

- NetWare is a discontinued computer network operating system developed by Novell, Inc. It initially used cooperative multitasking to run various services on a personal computer, using the IPX network protocol.

NOTES

4.8 KEY WORDS

- **Multi-tasking:** A type of OS that permits multiple programs to be run simultaneously by the same computer. A user of the computer can simultaneously play games while a Word document is being printed (The user is simultaneously working with two different applications–Word and Games). Operating systems supporting multi-tasking include UNIX and the Windows range.
- **Multi-threading:** A form of multi-tasking that permits multiple parts of a software program to be run simultaneously; for example, a user can perform a spell check on a Word document and simultaneously print another Word document. (User is working with two different components (Spell Check) and printing of the same application (Word). Operating systems supporting multitasking include UNIX and Windows.
- **Multi-processing:** This concerns the use of multiple processors (more than one CPU) to simultaneously execute multiple programs. The inclusion of multiple CPUs in a single computer system improves the performance to a large extent. Multiprocessing involves simultaneous processing by a computer system having multiple CPUs, whereas multitasking involves simultaneous processing by a computer system with a single CPU. Operating systems supporting multiprocessing include UNIX and Windows NT.
- **Single user:** This category of OS does not permit multiple users to use the computer and run programs at the same time. This assumes that at any given time only one user uses the system and runs only one program, i.e., it does not allow two users to concurrently work on the same program. MS DOS is an example.
- **Multi-user:** This class of OS permits multiple users to use the computer and run programs at the same time, e.g., UNIX, Linux, and Windows NT.
- **Linkage Editor :** It is an editor program that establishes a single module from many modules by resolving cross-references among the modules.
- **Utility programs**: These programs run on the operating system in order to carry out various user related commands to manage the software linking the user and the operating system. In MS-DOS, for example, the utility programs are: FDISK, FORMAT, ATTRIB, BACKUP, FIND, and others.

4.9 SELF ASSESSMENT QUESTIONS AND EXERCISES

NOTES

Short-Answer Questions

1. State the limitations of UNIX operating system.
2. Define resource management in OS.
3. What are the operating system services?
4. Explain the significance of Windows software.
5. What are the features of Linux?
6. List the important factors about Windows NT.
7. Differentiate between LAN and WAN giving appropriate examples.
8. Why is NetWare different from other operating system?

Long-Answer Questions

1. Briefly discuss the significance, uses, functions and components of operating system that is used in computer system. Support your answer giving appropriate examples and diagrams.
2. Write about the various types of operating system in detail.
3. Describe about the architecture of UNIX and its features.
4. Discuss about the concept of file in MS-DOS in detail.
5. Explain briefly about the various versions of Windows.
6. Briefly discuss the overview of networking in the context of LAN and WAN services.
7. Elaborate the history of NetWare operating system along with its features.

4.10 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

NOTES

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

BLOCK - II
BASIC CONCEPTS AND OPERATING SYSTEMS

NOTES

UNIT 5 BASIC CONCEPTS OF INTERNET

Structure

5.0 INTRODUCTION

NOTES

The Internet is the global system of interconnected computer networks that uses the Internet protocol suite (TCP/IP) to communicate between networks and devices. It is a network of networks that consists of private, public, academic, business, and government networks of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies. The Internet carries a vast range of information resources and services, such as the inter-linked hypertext documents and applications of the World Wide Web (WWW), electronic mail, telephony, and file sharing.

Principally, the Internet is basically a global network of computing resources, i.e., a physical collection of routers and circuits as a set of shared resources. It is a typical network of networks based on the TCP/IP communications protocol, i.e., Transmission Control Protocol (TCP) and Internet Protocol (IP). Consequently, the Internet is the collection of the many different systems and protocols. The TCP/IP communication protocols permits computers to transmit data on the Internet, therefore, with these protocols, virtually all computers can communicate with each other. Domain Name System (DNS) is an Internet address which has four fields with numbers that are separated by periods or dots, for example '198.105.232.4.', this type of address is known as an IP address.

The Internet service World Wide Web (WWW or W3), developed in 1989, is actually one of those different protocols. As the name implies, it permits resources to be linked with great ease in an almost unified and integrated manner. Basically, the WWW contains a vast collection of linked multimedia pages that is ever-changing. There are several basic components of the Web that permit the users to communicate with each other. To search on the Internet, a 'Web Browser' is required, which is a software program that permits the users to view 'Web Pages' and navigate through the Internet. Microsoft's Internet Explorer (IE) and Chrome is probably the most commonly used web browsers. A web page is a single page of information on the World Wide Web. A 'Website' refers to a group of web pages identified by a single domain. A 'Homepage' is referred as the first or front page of a website. A URL or 'Uniform Resource Locator' is the unique address of each web page. Each URL has several parts that appear in a specific order.

A 'Search Engine' can be defined as a website that uniquely searches the World Wide Web for specific keywords, which is entered by the user into a search field. The search engine then displays a list of web pages that are somehow approximately related to the keywords that was entered by the user for search. The user then 'Clicks' on the links to any of the displayed list of web pages that are significant.

MODEM is the shortened word for MOdulator-DEModulator. Basically, a MODEM is an important equipment of a computer, and its purpose is to relay

NOTES

or transmit information either using a telephone or a cable line. Modems can be attached to computers either externally or placed internally.

Electronic mail (email or e-mail) is a method of exchanging messages electronically through 'mail messages' between people using electronic devices. Ray Tomlinson is credited as the inventor of email; in 1971, he developed the first system able to send mail between users on different hosts across the ARPANET, using the @ sign to link the user name with a destination server.

In this unit, you will study about the basics of basic concepts of the Internet, Modem, e-mail, Websites, address, domain, protocols, types of accounts, search engines, browsing Web, Telnet, Usenet and AOL.

5.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the basic concepts of the Internet
- Define what MODEM is
- Analyse the importance of e-mail
- Explain the significance of Websites and address for the Internet
- Elaborate on domain, protocols and types of accounts on the Internet
- Discuss about the functioning of search engines and methods of Web browsing
- Define the features of Telnet, Usenet and AOL

5.2 BASIC CONCEPTS OF INTERNET

The Internet, World Wide Web and Information Super Highway are terms which have the lives of millions of people all over the world. The widespread impact of Internet across the globe could not be possible without the development of Transmission Control Protocol/Internet Protocol (TCP/IP). This protocol suite is developed specifically for the Internet. The Information Technology revolution could not have been achieved without this boundless chain of networks. It has become a fundamental part of the lives of millions of people all over the world. All the aforesaid services, provide us the necessary backbone for information sharing in organizations and within common interest groups. That information may be in several forms. It can be notes and documents, data to be processed by another computer, files sent to colleagues, and even more exotic forms of data.

During late 60s and 70s, organizations were inundated with many different LAN and WAN technologies such as packet switching technology, collision-detection local area networks, hierarchical enterprise networks, and many other excellent technologies. The major drawbacks of all these technologies were that

NOTES

they could not communicate with each other without expensive deployment of communications devices. These were not only expensive but also put users at the mercy of the monopoly of the vendor they were dealing with. Consequently, multiple networking models were available as a result of the research and development efforts made by many interest groups. This paved the way for development of another aspect of networking known as protocol layering. This allows applications to communicate with each other. A complete range of architectural models were proposed and implemented by various research teams and computer manufacturers. The result of this know-how is that today any group of users can find a physical network and an architectural model suitable for their specific needs. This includes cheap asynchronous lines with no other error recovery than a bit-per-bit parity function, through full-function wide area networks (public or private) with reliable protocols such as public packet switching networks or private SNA networks, to high-speed but limited-distance local area networks.

It is now evident that organizations or users are using different network technologies to connect computers over the network. The desire of sharing more and more information among homogeneous or heterogeneous interest groups motivated the researcher to device a technology whereby one group of users could extend its information system to another group who had a different network technology and different network protocols. This necessity was recognized in early 70s by a group of researchers in the United States of America (USA) who hit upon a new principle popularly known as Internetworking. Other organizations also became involved in this area of interconnecting networks, such as ITU-T (formerly CCITT) and ISO. All were trying to define a set of protocols, layered in a well-defined suite, so that applications would be able to communicate with each other, regardless of the underlying network technology and the operating systems where those applications run.

Internetworks

The availability of different operating systems, hardware platforms and the geographical dispersion of computing resources necessitated the need of networking in such a manner that computers of all sizes could communicate with each other, regardless of the vendor, the operating system, the hardware platform, or geographical proximity. Therefore, we may say that *Internetworking* is a scheme for interconnecting multiple networks of dissimilar technologies. To interconnect multiple networks of dissimilar technologies use both additional hardware and software. This additional hardware is positioned between networks and software on each attached computer. This system of interconnected networks is called an *Internetwork* or an *Internet*.

To develop standards for Internetworking, Defense Advanced Research Projects Agency (DARPA) funded research projects. ARPANET, a project of DARPA, introduced the world of networking with protocol suite concepts such as layering, well before ISO's initiative in this direction. DARPA continued its

research for an Internetworking protocol suite. This may be seen in the early NCP (Network Control Program) host-to-host protocol to the TCP/IP protocol suite, which took its current form around 1978. DARPA was well known for its pioneering of packet switching over radio networks and satellite channels and ARPANET was declared an operational network with responsibility of administering it to Defense Communications Agency (DCA) in 1975. TCP/IP had not yet been developed.

ARPANET was basically a network based on leased lines connected by special switching nodes, known as Internet Message Processors (IMP). Many researchers were involved in TCP/IP research by 1979. This motivated DARPA to form an informal committee to coordinate and guide the design of the communication protocols and architecture. The committee was called the Internet Control and Configuration Board (ICCB).

The first real implementation of the Internet was when DARPA converted the machines of its research network ARPANET to use the new TCP/IP protocols. After this transition which started in 1980 and finished in 1983, DARPA demanded that all computers willing to connect to its ARPANET must use TCP/IP. The US military adopted TCP/IP as standard protocol in 1983 and recommended that all networks connected to the ARPANET conform to the new standards.

The success of ARPANET was more than the expectations of its own founders and TCP/IP Internetworking became widespread. As a result, new Wide Area Networks (WAN) were created in the USA and connected to ARPANET using TCP/IP protocol. In turn, other networks in the rest of the world, not necessarily based on the TCP/IP protocols, were added to the set of interconnected networks. Computing facilities all over North America, Europe, Japan, and other parts of the world are currently connected to the Internet via their own sub-networks, constituting the world's largest network. In 1990, ARPANET was eliminated, and the Internet was declared as the formal global network.

DARPA also funded a project to develop TCP/IP protocols for Berkeley UNIX on the VAX and to distribute the developed codes free of charge with their UNIX operating system. The first release of the Berkeley Software Distribution (BSD) to include the TCP/IP protocol set was made available in 1983 (4.2BSD). This led to the spread of TCP/IP among universities and research centers and has become the standard communications subsystem for all UNIX connectivity. There are many updated versions of BSD code available. These are 4.3BSD (1986), 4.3BSD Tahoe (1988), 4.3BSD Reno (1990) and 4.4BSD (1993).

Some examples of the different networks that have played key roles in this development are described below:

The Internet

The word 'Internet' is a short form of a complete word Internetwork or interconnected network. Therefore, it can be said that the Internet is not a single

NOTES

network, but a collection of networks. The commonality between them in order to communicate with each other is TCP/IP. The Internet consists of the following groups of networks:

- **Backbones:** These are large networks that exist primarily to interconnect other networks. Some examples of backbones are NSFNET in the USA, EBONE in Europe and large commercial backbones.
- **Regional Networks:** These connect, for example, universities and colleges. ERNET (Education and Research Network) is an example in the Indian context.
- **Commercial Networks:** They provide access to the backbones to subscribers, and networks owned by commercial organizations for internal use and also have connections to the Internet. Mainly, Internet Service Providers come into this category.
- **Local Networks:** These are campus-wide university networks.

The networks connect users to the Internet using special devices that are called gateways or routers. These devices provide connection and protocol conversion of dissimilar networks to the Internet. Gateways or routers are responsible for routing data around the global network until they reach their ultimate destination as shown in Figure 5.1. The delivery of data to its final destination takes place based on some routing table maintained by router or gateways. These are mentioned at various places in this book as these are the fundamental devices to connect similar or dissimilar networks together.

Over time, TCP/IP defined several protocol sets for the exchange of routing information. Each set pertains to a different historic phase in the evolution of architecture of the Internet backbone.

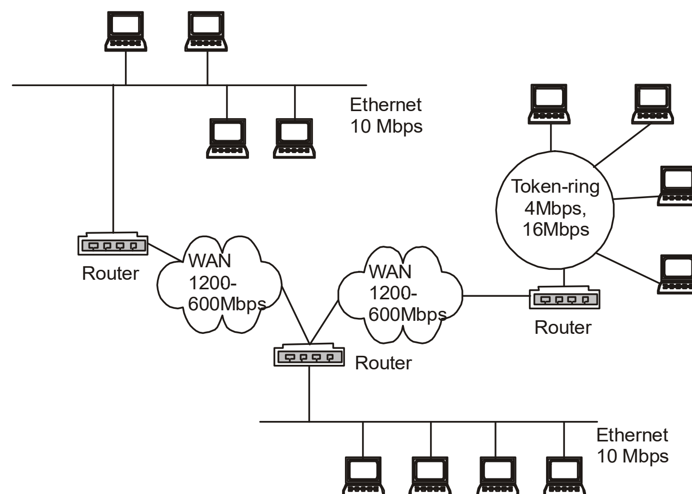


Fig. 5.1 *Local Area Networks Connected to the Internet via Gateways or Routers*

NOTES

ARPANET

ARPANET was built by DARPA as described earlier. This initiated the packet switching technology in the world of networking and therefore is sometimes referred to as the "grand-daddy of packet networks". The ARPANET was established in the late 60s for Department of Defense to accommodate research equipment on packet switching technology besides allowing resource sharing for the Department's contractors. This network includes research centres, some military bases and government locations. It soon became popular with researchers for collaboration through electronic mail and other services. ARPANET marks the beginning of Internet.

ARPANET provided interconnection of various Packet-Switching Nodes (PSN) located across continental USA and Western Europe using 56 Kbps leased lines. ARPANET provided connection to minicomputers running a protocol known as 1822 (after the number of a report describing it) and dedicated it to the packet-switching task. Each PSN had at least two connections to other PSNs (to allow alternate routing in case of circuit failure) and up to 22 ports for user computer connections. Later on, DARPA replaced the 1822 packet switching technology with the CCITT X.25 standard. The increase in data traffic made 56 Kbps capacity of the lines insufficient. ARPANET has now been replaced with new technologies as backbone for the research side of the connected Internet.

5.2.1 Domain Name System (DNS)

The Domain Name System (DNS) is a distributed database that provides e-mail routing information. It is used by TCP/IP protocols. These protocols map between IP addresses and hostnames. The domain name extension type is tagged with distributed database because no single website across a network can access all the information. Each site runs a server program for sites and maintains its own database of information. Therefore, collecting information is possible through distributed method that creates a mechanism for clients and servers to communicate with each other.

The Domain Name Extension (DNE) is defined as the complete address of hosting services provided on the sites. In the beginning, the Internet configuration used numeric IP address, which was a very cumbersome task. To overcome this problem, symbolic host names came into existence. For instance, initially, it was typed as TELNET 10.12.7.14 but nowadays, TELNET MyHost command is issued. With this command, the mapping between machine names and IP addresses has also become centralized and coordinated. A name space is organized in two ways—either hierarchical or flat. It basically maps each address to a unique specified name. The address mappings of host names are maintained by Network Information Centre (NIC) which maintains a single file known as the flat name space. Hierarchical name space involves the nature and name of the organization but in this case, the controlling authority is decentralized. The design of domain

NOTES

name space is hierarchical. It follows inverted tree structure, that is, from the root to the top. It can have up to 128 levels that start from level 0, that is, the root level and may go up to level 127. A label can have a maximum of 63 characters. The root level contains a Null string. Each node in the reversed tree keeps a domain name. A fully qualified domain name contains a sequence of labels separated by dots. The first part defines the nature of an organization, the second part signifies the name of the organization and the third part refers to the department of the organization. (see Figure 5.2). The authority to assign and control the name spaces can be decentralized. For example, myHost.myDept.myDiv.myCorp.com

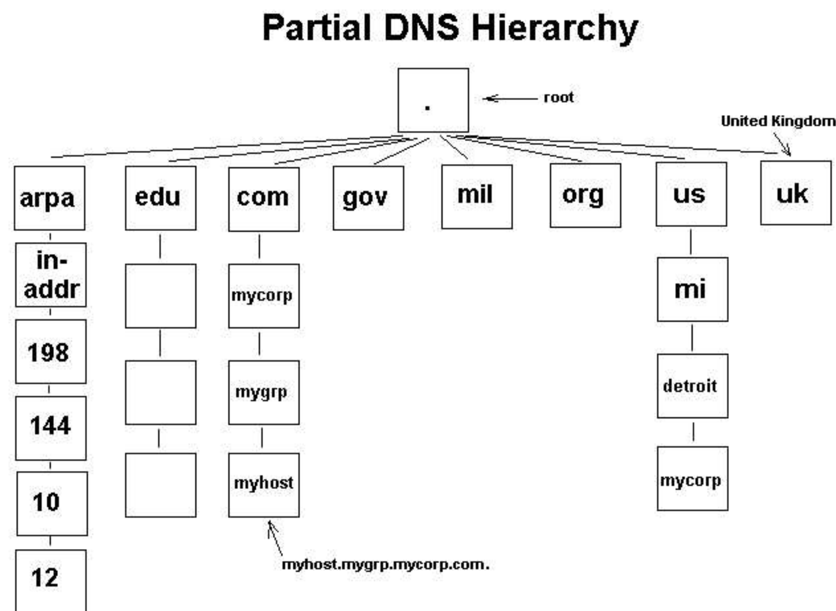


Fig. 5.2 *Domain Name Extension Hierarchy*

The domain name space is hierarchical in design. The names are defined in an inverted-tree structure with the root at the top. The tree can have 128 levels, that is, level 0 (root) to level 127. Each node in the tree has a label, which is a string with maximum of 63 characters. The root label is a null string. DNS requires that the children of a node have different labels. Each node in the tree has a domain name. A full domain name is a sequence of labels separated by dots. The domain names are always read from the node to the root. Table 5.1 shows the top-level domains that are assigned in the United States.

If the domain name ends in a dot, it shows that the name is not complete. This is known as AND. It is referred to as a Fully Qualified Domain Name (FQDN) or an absolute domain name. For example,

myDept.myDiv.myCorp.com.

If the domain name does not end in a dot, then it means that it is complete. This is called Partially Qualified Domain Name (PQDN). For example,

myDept.myDiv.myCorp

NOTES

The nature of DNS is distributed where symbolic names are grouped into zones of authority. These zones contain a database of symbolic names along with IP addresses. Each zone is a part of sub-tree of hierarchical structure. The names are administered independently within the zones and can be assigned to other zones. The name server includes authority over the zones. Figure 5.3 shows that the domain name can be searched by its extension type.

Search for your domain name

www. dev (e.g yourcompany)

.com .net .org .info .name
.us .biz .ca .tv .eu
.ro .ws .co.uk .de

Search

Fig. 5.3 *Searching Domain Name by its Extension Type*

Now we have two types of IP addresses in the form of decimal numbers and text for the same host. You know that list of all IP addresses are maintained centrally by ICANN in the form of distributed database directory. There are several distributed servers, which maintain this list of IP addresses. The reasons behind the distributed server are very logical and simple. It helps in disaster management and in diverting the load of traffic in the form of requests from clients to other DNS servers located at different sites. DNS server maintains database in both the textual and the decimal notation form. For example, DNS server maintains the address of the google site as www.google.com and 216.23.9.53.99. In this manner, DNS is used to provide host-to-IP address mapping of remote hosts to the local hosts and vice versa. It is now conspicuous that the DNS maintains a distributed database to map between hostnames and IP addresses. Whenever a client requests a service from a site, both the sites run DNS protocol to access the distributed database which is nothing but Domain Name Systems. Therefore, the DNS provides the protocol, which allows the clients and servers to communicate with each other. DNS also helps the system use a resolver, which resolves the host name to that IP address, which can be understood by the server.

Now let us talk about how DNS is able to quickly translate the text of the IP addresses, from a directory of billions of such addresses and that too within a fraction of seconds.. This could be made possible by using Domain concepts, which use hierarchical arrangements of text addresses translation.

As illustrated in the Figure 5.4 you can see that at the top level is the root server, which has a null label. Below this is another level domain, namely com, edu, int which are grouped together. Below this different sub domains or groups have been created. Table 5.1 depicts some commonly appearing domain names with their respective sites. The DNS can accommodate almost all kinds of organizations by allowing each group to choose between geographical or organizational naming hierarchies.

NOTES

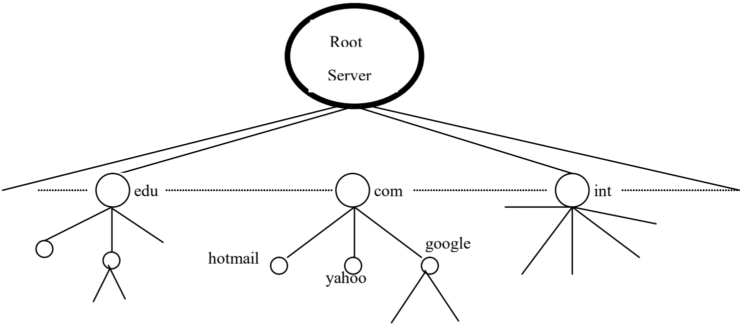


Fig. 5.4 *DNS Hierarchy*

Table 5.1 *Internet Domains*

Domain	Indicative Site
Com	Commercial institute
Edu	Educational institute
Org	Non-profit organization
Net	Network service provider
Gov	Government department
Mil	Military
Biz	Business
Country code	For example, in for India, us for USA, au for Australia, jp for Japan and so on

As we know that the servers maintaining addresses are disseminated and have locations throughout the world. Now the questions arises that how text addresses are organized in hierarchical arrangement. Now consider the Figure 5.4 and Table 5.1 given above. The hierarchy is represented through zones and each zone depicts a hierarchy of one or more nodes without any overlapping. Each zone is represented by a server and undoubtedly with one backup server. The Root server, as shown in Figure 5.4, is the only one which is just indicative. There may be several root servers at several locations in the world. Each root is aware of the location of each DNS server of a specific domain.

The process is now very simple to understand. When you need to connect with a particular site, you first send your request to your local host. If your local host provides the translation, your request is completed. If not, your local host then sends your request one level above in the hierarchy. If the server at one level above is able to handle the same, you get your intended website at your desktop through your local server. If not, then the server at one level above your local server either sends your request again to another server or informs your local server that your request has failed and eventually gives you the address of another server to process your request. This process continues until a server, which knows the address, is found. Otherwise, the request is filtered up to the root server.

Depending upon the domain address, the root server forwards the request to one of the domain servers represented at the next level of hierarchy. This process continues until the information of text address is returned to the Root server and then back to your local server.

5.3 INTERNET SERVICES

The Internet is known as 'the Network of Networks'. It is like a phone system that connects almost anywhere around the world. It exchanges information and acts as global link between small regional networks. Internet services offer a gateway to a myriad of online databases, library catalogues and collections, and software and document archives, in addition to frequently used store-and-forward services, such as UserNet News and e-mail. The widely used Internet services are as follows:

E-Mail

E-mail is the prime Internet service that facilitates services to people or users across the world. Full Internet connectivity is not required for this. For example, an electronic address provides these services to FTP sites through which mail can be exchanged. Other Internet services, such as IP address resolver, Archie lookup, WHOIS service is done via e-mail.

The header and body of the message make an e-mail message. The header contains the information where the message is to be sent and the complete path for reaching the destination, date and return path. The body of the message is the actual message that has to be sent. The syntax of an e-mail address is user@ subdomain.subdomain.domain, e.g., **abc@gmail.com**. A service provider must be connected with leased line, dial-up or connection with any network for sending e-mail.

File Transfer Protocol (FTP)

FTP is also prime Internet service that acts as protocol and transfers files over TCP/IP network (Internet, UNIT, etc.). Once HTML page is developed on a local machine for a website, it is first uploaded to the Web server through FTP. Local machine is the machine on which you are initially logged into. It includes functions to log on to the network, gives a list of directories and copies files. FTP transfer is possible by entering URL preceded with ftp:// within address bar of a web browser. The FTP operations can be performed by issuing FTP commands at the command prompt or by using FTP utility running under a graphical user interface on Windows OS. FTP tasks can be performed through a browser. For example, type an IE address bar URL as ftp:// to get ftp services. For example, ftp://YourLoginName@IPaddress.

The required steps used in connecting with FTP operations are as follows:

NOTES

The local machine is connected with remote machine by typing 'ftp machinename'. The machinename is the full name, written as aaa.cs.state.edu, of the remote machine to which the local machine is to be connected. Basically, the machine name is the remote machine's full name. If the machine name is not available other option is taken as to type the 'ftp machinenumber' that demands the Net address of the remote machine, e.g., 129.15.0.11. The FTP responds to the users to enter their loginname and password. The anonymous ftp is used widely these days. Many computer systems provide this facility so that you can access the information of specific machine without creating an account on that particular machine. These types of services are provided by anonymous FTP. You need not to be registered user of the system. The anonymous FTP server contains relevant software, documents and files used to configure networks, graphics, images, songs, lyrics and other useful information. An electronic e-mail can be archived through the anonymous FTP. The ready information is stored in machines for any user across the Net who wants to get the required information.

Telnet

Telnet is used to connect remote network computers. It is the Internet service that executes commands on remote host as if you are going to log in locally. For this, the machine name and valid user name are required to be connected. The commands that are issued on telnet are as follows:

***Telnet Hostname*:** A connection to the host name is opened by this command. For example, issuing the command as 'telnet abc.maths.edu' with that machine which keeps the required information of abc.maths.edu site can connect you.

***Telnet Address*:** It gives the IP address of the connected host.

Archie

If some programs are installed in a system unit and you want to know the availability of the program on the Internet, you can get to know the machine along with such programs via Archie. Basically, Archie is a program that searches files anywhere on the Net by filename. This facility is maintained by a database with the Internet sites accessible via anonymous FTP. The following table 5.2 shows the various types of Archie servers:

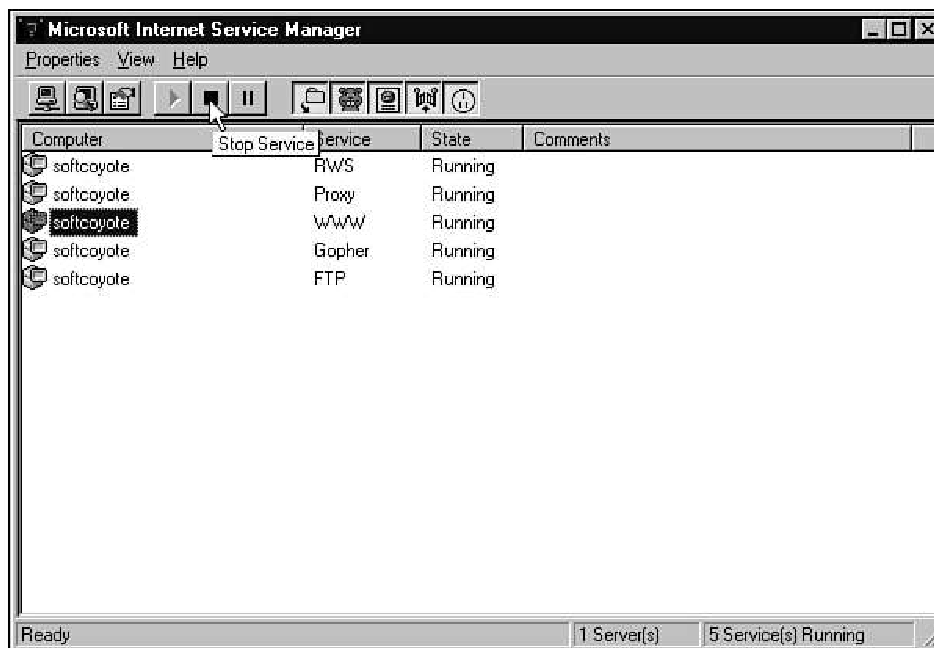
Table 5.2 *Various Types of Archie Servers*

Name	IP Address
archie.rutgers.edu	128.6.18.15
archie.cs.mcgill.ca	132.206.51.250
archie.funet.fi	128.214.6.102
archie.rediris.es	130.206.1.2
archie.sura.net	192.239.16.13
archie.doc.ic.ac.uk	146.169.16.11

The Archie server can be accessed via Telnet, for example, 'Telnet archie.rutgers.edu'. For getting Archie server login to 'Archie'. It requires no password. You can look for files by its full name. For this, either 'set type exact' syntax is used or you can use 'set type sub' syntax. The 'set type sub' syntax is used if the required name of the file is known. The 'find file-name' syntax is also used to find the required file name.

Gopher

The Gopher protocol supports client–server software that searches files on the Internet. A Gopher client is required for validating and testing of Gopher publishing service. For example, WS Gopher 1.2 is available on the Internet as shareware. The server based text files are hierarchically organized and viewed by end-users. These end-users access the server by using Gopher applications of remote computers. Gopher browsers initially display the text-based files. Most of the files along with database are available on Gopher that converts HTTP compatible formats and makes them available on the net.



In the preceding screen, the Internet service manager displays the services that are installed on the server to which the Internet service manager is attached.

Finger

Finger service gives information about users, for example, username, person's first name and last name, information about recently logged in and also where they logged in. But the users must enter the required information where they get registration for particular e-mail services. Finger is also used to get a list of users who are currently logged into the host. In fact, the Finger program accepts input

as an e-mail address that returns information of user. In some systems, Finger gives the information about the currently logged on users.

World Wide Web (WWW)

NOTES

WWW provides hypertext access to documents located anywhere on the Internet. It is a very successful distributed information system. It is basically client–server data transfer protocol that communicates via application level protocol. Its structural components are clients–browsers, servers and caches. The Internet and semantic components include HyperText Transfer Protocol (HTTP), HyperText Markup Language (HTML) eXtensible Markup Language (XML) and Uniform Resource Identifiers (URIs). The clients who get various sites requested to the server via HTTP determine the structure of WWW. Then web pages constructs HTML consists of graphics and sound embedded files. For running the complete system, TCP/IP, DNS networking protocols are required.

The reason behind the evolution of Java programming language is to develop distributed application. Distributed application means many CPUs are inter-connected through different network topology so that each CPU can communicate with one another. Java introduced the remote method invocation technique to implement distributed application. The **java.net** package provides classes and methods to develop networking-applications through different network protocols.

A group of computers connected by cable to share information is popularly known as network. A network is a set of computers and peripherals that are physically connected. Networking enables sharing of resources and communication. Java applets can be downloaded from a website. This is one of the main attractions of Java. Networking in Java is possible through the use of **java.net** package. The classes within this package encapsulate the socket model developed by Berkeley software division. The network requires some components, such as:

- Server
- Client
- Peer
- Protocol
- Physical Media
- Physical Devices

Servers provide services to the client. If a server provides application services, then it is treated as an application server.

The **client** accesses services from the server. Peer is a computer that works as a server as well as a client.

Clients

A computer, which requests for some service from another computer, is called a **client**. The one that processes the request is called a **server**. A **server** waits till

one of its **clients** makes a request. It can accept multiple connections at a time to the same port number. Multithreading is used to serve multiple users at the same time.

NOTES

Check Your Progress

1. What do you mean by Internetworking? How it works?
2. What is Internet?
3. Define ARPANET?
4. Define the term DNS.
5. What is an e-mail? How an e-mail is composed and sent?
6. How are FTP operations performed?
7. Why is Telnet used?
8. What is Gopher? Which files it supports?
9. What is Finger service? What it its use?

5.4 ELECTRONIC MAIL

Electronic mail is one of the most popular network services. The use of e-mail is considered the foremost reason behind the popularity of Internet. The proliferation of cyber cafés can be attributed to e-mail or World Wide Web. E-mail provides an efficient and fast means of communication with relatives, friends or colleagues throughout the world. With the help of email, one can not only communicate with myriad people at a time but can also receive and send files and other information within a fraction of seconds. The biggest advantage of email is that the intended receiver of the message does not require to be present at their desktop at the time of receiving of the message.

5.4.1 Definition

The term email connotes the basic communication facility provided by the Internet to its users to send and receive messages in any part of the world. It is considered one of the most popular applications of the Internet and is accounted for 90% of net traffic. Email facilitates sending of messages in the form of a text, audio and video or even a combination of these types. When a message is sent from the source user, it reaches the recipient's mail box. The email message received by the recipient can be opened, discarded, edited, saved, responded back to or can even be forwarded to some other recipient. Email messages are delivered instantly after the transmission. An email can be sent by connecting to the network from any location. An Internet connection usually requires a telephone line, a modem and a computer. Wireless connections have also become popular means of getting connected to the Internet. This job is accomplished by the Simple Mail Transfer Protocol (SMTP) running over TCP/IP.

5.4.2 Uses of Email

NOTES

Email provides several features that are useful in day-to-day life. It is an efficient and cost-effective way of communication across the world. With the help of email, one can send common letters or circulars to several recipients. The email messages are delivered instantly, even if they are sent to remote locations worldwide. Thus, it saves time as well as money. Whereas the postal messages are time consuming. Email also provides an address book facility which keeps a record of the email addresses. This saves the user from the predicament of remembering the addresses of the recipients. In addition, a lot of time, energy and money is saved as the user creates a mailing list with a group name, so that a letter or a circular can be transmitted by just typing the name of the particular group. Another advantage of using email is that provided the email address typed in is correct, it enables the sender to know immediately whether the message has been delivered to the recipient,. In case the message is not delivered, the sender will receive a return email message to inform him about the failure of the particular message. Email goes beyond all time zones and barriers.

Email also provides the user with a facility of attachment which allows the user to attach any file created in any application such as word processors, spread sheet or power point presentations. For example, if the total amount of outstanding against a client is computed in a spreadsheet, the client may be informed through a letter in email along with an attachment showing his outstanding amount in the spreadsheet. The primary advantages of email can be summarized in the following:

- It conducts paperless communication of messages quickly.
- It ensures simultaneous transmission of messages to several users. The messages may comprise of pictures, video, film clips, text, animation or even a combination of them. Voice and audio messages can also be transmitted this way.
- The email messages can also be printed, prioritized, forwarded and stored.
- Public bulletin boards can be created in which every member of the organization can post and view messages. This can also be accomplished in the case of shared text messages and application files used widely across computer platforms.
- It allows delivery and receiving of faxes and meetings can also be scheduled through email.

5.4.3 Opening an E-Mail Account

Opening an E-mail account is not an issue. Now-a-days, all subscribers get facility to open an email account free of cost. A number of web services like Google, Hotmail, Yahoo, etc. are readily available to register a user to open an email account and access it from anywhere in the world. However, to avail this facility, the user

NOTES

should have access to a computer and an Internet connection. In addition to these web services, organizations or ISPs also provide web interfaces to enable the users to open their email accounts, though by charging them. In this case, the organization or ISP possesses the personal record of the users and based on their personal records and their relationship with them, they open their email accounts and equip them with an email address. The email addresses comprise of email ids meant for individual users, which could be their first name or a combination of their name and surname or their date of birth, etc. along with the URL of the organization. For example, in sanjay0203@teraclean.com, sanjay 0203 signifies email id consisting of the name and birth date and month, whereas teraclean.com indicates the URL of the organization.

In case of universally available web services like Google, Hotmail or Yahoo, the user needs to open the website of the respective Web service by typing its corresponding URL in the Web browser. For example, if the user wants to open an account in the Google Web service, then he needs to key the Web address of Google, that is, www.google.com. Once the Website of Google opens, the user needs to click on the Gmail service of the Google. The Gmail interface provides the facility for opening of a new account, for which it provides a registration form to be filled up by the user. In accordance with the procedure, the user mentions his personal information, email id and password in the form. Thereafter, he gets registered and obtains an email address. This process of creating an email account is described as follows:

Type the URL "http://www.google.com" in the address bar of a Web browser such as an Internet Explorer, to visit the Google homepage as shown in Figure 5.5.

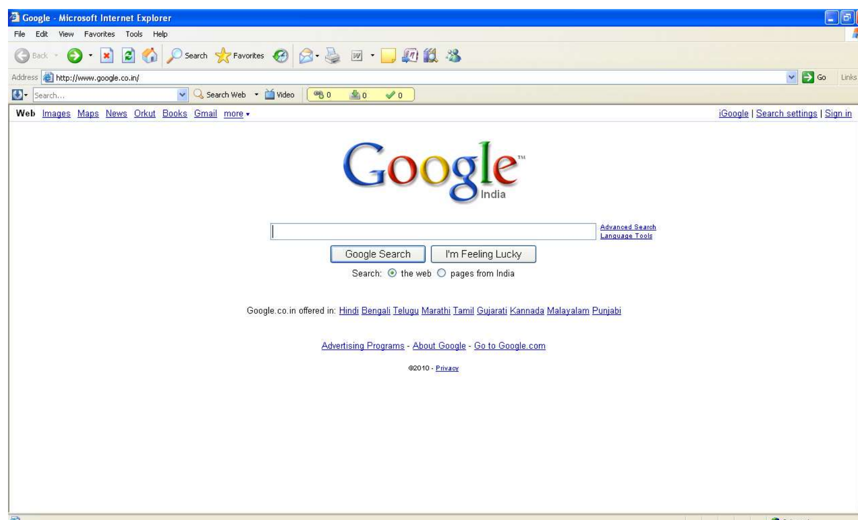


***Fig. 5.5** Google Homepage*

The page displayed shows an icon namely Gmail as shown in Figure 5.5. Once you click on the Gmail icon, it navigates you to another webpage as shown in Figure 5.6. If you have an existing account with Gmail, you can type in your

Email ID and your password to log on to your account. If you are accessing the Gmail for the first time, then you need to create an account for yourself. The procedure for the same is as follows:

NOTES

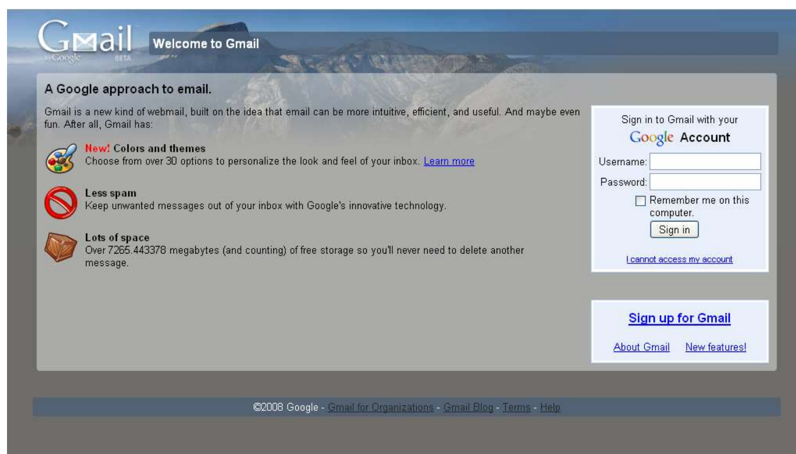


Fig. 5.6 Sign-In Page

Click on the "Sign Up" icon as illustrated in Figure 5.6. This will lead you to another webpage that contains the registration form as shown in Figure 5.7. Now you are required to fill the form that asks for your personal details along with your user ID and password to open a new email account for you.



Fig. 5.7 Registration Page

NOTES

Once you are through with the registration process, after accepting the terms and conditions, you become a member and thereafter you are able to use your email account to send and receive emails. Now all you need is to remember your User ID and your password for future use of your email account. In other words, it means that whenever you need to log on to your newly created account, you need to simply type your User ID and your password.

5.4.4 Reading and Writing E-Mail

Email is extensively used by people across the world. The procedure of reading and writing an email is not a very sophisticated one. The steps involved are as follows:

Reading an E-Mail Message

The email account can be accessed at anytime and from anywhere by logging on to the particular email account, as mentioned earlier. To read or write an email, you need to perform the following steps:

1. Type the URL "http://www.google.com" in the address bar of a Web browser.
2. Enter your User ID and the password as shown in Figure 5.8.

Fig. 5.8 *Sign-In Page*

Once you have signed in successfully, you can access your email account as shown in Figure 5.9.

NOTES

Add a mail account

Enter the mail settings for you@yourdomain.com. Learn more

Email address: you@yourdomain.com
Username: you@yourdomain.com
Password: •••••••••••
POP Server: mail.yourdomain.com Port: 995
Leave a copy of retrieved message on the server. Learn more
Always use a secure connection (SSL) when retrieving mail. Learn more
Label incoming messages: you@yourdomain.com
Archive incoming messages (Skip the Inbox)
Cancel « Back Add Account »

Fig. 5.9 *E-Mail Account*

Clicking on the Inbox icon lets you open your Inbox. The Inbox folder contains all your previous email messages and also enables you to read the new ones. You also have an option of deleting the previous messages or transferring them to some other folders also. An email message in the Inbox can be read by clicking on the email subject or any other clickable item therein. This displays the contents of the message to be read and allows you to take appropriate action accordingly. Some email messages are delivered along with attachments. Attachments may comprise of textual messages, graphics, pictures, videos, sounds or a combination of these types.

The email message depicts an attachment button within the message itself, which on being clicked enables you to either open the attachment in relevant applications or save it on your computer to be opened separately.

Writing an E-Mail Message

As mentioned earlier, email account can be accessed by logging on to email account. To write an email, you need to perform the following steps:

The Compose option on the left hand side of the screen enables you to write an email message. Attachments can also be appended along with the email messages wherever they are required. On selection of the compose option, a screen as shown in Figure 6 will appear. The following steps are to be followed for writing and sending an email message:

1. **To:** It is a field in which the valid email address of the recipient like User ID@domain.com is typed in, so that the message can be delivered correctly. In case of multiple recipients, email address of each recipient is typed in the same box separated by comas.

NOTES

2. **Cc:** It signifies the email address/(s) of the recipient/(s) to whom a carbon copy of the message is to be transmitted. The recipient/(s) specified in To field also receives the email address/(s) of the recipients in their messages indicating that email address/(s) in the Cc field also receive/(s) the same message.
3. **Bcc:** It denotes the email address/(s) of the recipient/(s) to whom a copy of the message is transmitted. However, in this case, the recipient/(s) in both To and Cc field remain oblivious of the other email addresses, to which the message is sent. Bcc stands for blind carbon copy.
4. **Subject:** This box enables the sender to write the subject of the message, so that recipient/(s) on receiving the message, could have a clear idea of what the email message is about.
5. **Message Box:** It is the field in which you type your message which is to be transmitted.

An attachment can also be appended to the email message before sending it. There exists an Attachment button within the compose mail box. On clicking on the Attachment button, you are asked to provide the location of the desired file to be attached. You then click on the Browse button which enables you to select the desired file from your computer. Finally, clicking on the Attach or OK button attaches the document along with your email message.

Your message with or without attachment is now ready to be transmitted. Now you need to follow the following steps:

1. If you want to postpone transmitting of your message, you have another option called Draft in which you can save your message to be transmitted later. The message saved in the Draft can also be modified before transmission. The Draft webpage provides you a Send button. On clicking on it, your message is transmitted and a copy of the message is saved in your Sent mailbox, provided the send and save option has been set.
2. If you do not want to postpone the transmission of your message, then just click on the Send button. On clicking on it, your message will be transmitted and a copy of the message will be saved in your Sent mailbox.

Check Your Progress

10. Define the primary advantages of e-mail.
11. How Gmail helps in opening an email account?
12. What does an Inbox folder contains?

5.5 EQUIPMENTS REQUIRED FOR AN INTERNET CONNECTION

NOTES

Surfing the Internet is quite similar to scuba diving, with regard to the sophisticated equipments deployed to access Internet. Just as we need certain equipments to dive in the deep ocean for scuba diving, we require adequate paraphernalia to successfully plunge into the huge ocean of interconnected computers and networks.

Now-a-days, Internet access necessitates a broadband connection, which is, a high data rate Internet access. The dial-up access deploys a 56K dial up modem, which uses a dedicated telephone line and is limited to the bit rate of less than 56 Kbps. In contrast to this is the broadband technology, which provides more than double the dial up bit rate and that too without intervening with the telephone use. In other words, it means that Internet access and voice call can be carried out simultaneously. The broadband connections are characterized by various minimum bandwidths ranging from 64 Kbps up to 2.0 Mbps. Some standards define the broadband connection as having download data transfer rates equal to or faster than 256 Kbps, whereas others define it as having data transmission speed exceeding 768 Kbps in either downstream or upstream direction. In general, any connection of 256 Kbps or greater comes under broadband Internet.

Certain equipments which are required to access the Internet are as follows: Amongst these, some of them are mandatory and some are optional:

- **Computer:** A computer which is used to browse the Internet may either be a personal computer with Pentium processor or a Macintosh. It should have enough power and memory concomitant with multimedia features. Though 128 MB RAM is sufficient to have access to Internet but 512 MB RAM or more is recommended. Now-a-days, devices like smart phones, mobile phones, Pocket PCs, etc., are also used to browse the Internet.
- **MODEM:** It stands for MOdulator/DEModulator. This may either be internally built in or externally connected. The MODEM is a device that converts data in binary code used by the computer, to an analog signal that can be transmitted over the telephone network and vice versa.

  With the help of telephone lines, millions of computers worldwide are connected with one another, either directly or indirectly. In order to connect with the Internet Service Provider (ISP), these connections require the regular dial up telephone lines or dedicated higher capacity telephone lines like leased lines, ISDN lines, etc.

- **Internet Account with a Service Provider:** An account with a service provider is essential to create a link between the user's computer and the Internet. A service provider, which is popularly referred to as ISP (Internet

Service Provider), signifies phone or cable companies that provide last mile connectivity. It may also refer to a cable line from the subscriber's home to his office and also to an exchange for long distance connectivity based on monthly or annual charges.

NOTES

- **Widely used Current Standard Broadband Technologies:** These technologies are DSLs, that is, Digital Subscriber Line (DSL) and cable modems. However, recent technologies like VDSL and optical fiber connections are also gradually becoming popular in providing Internet access in a much more cost-effective way than copper wire technology. Wi-Fi networks are also used to provide Internet connections. However, these are not served in the areas by cable or ADSL. The WiMAX has been gaining popularity with regard to mobile and stationary broadband access.
- **Internet Browser Software:** It is the software tool which enables a user to browse the Internet with the help of web addresses or URLs. A few of the widely used browsers are Internet Explorer version 7 or 8 (IE), Netscape, Mozilla Firefox, Chrome, AOL, Opera, etc.
- **Anti-Virus Software:** These are used to protect the user from the onslaught of the nasty programs that obtain access to the user's terminal when he is surfing the network or downloading contents from there. Some examples of anti-virus software are Symantec, Norton, McAfee, etc.
- **Email Software:** The email software may be chosen from the Outlook or Outlook Express. Google, Yahoo and Hotmail offer free web-mail for the same.
- **Plug-In Software:** It is considered an add-on to the user terminal. It enables the user to avail services like music, video, multimedia, etc. on the Internet. The most popular plug-in-softwares include Real Audio music player, Macromedia Flash Player, Windows Media Player, Apple Quick Time, Java Virtual Machine, etc.
- **Stereo Speakers, Microphone and Webcam:** These equipments enable the user to play sounds, videos, to conduct Internet telephoning and to send images to other users connected to the Internet.

5.6 WEB BROWSER

A browser is a software, which your computer uses to view WWW documents and access the Internet. The browser program residing in your computer provides you with the facilities like text formatting, hypertext links, images, sounds, motion and other features. Internet Explorer and Netscape are considered the most widely used browsers. Browsers have sub programs called plug-ins to handle the documents found on the Web. It may also have other plug-ins stored elsewhere in the computer.

NOTES

Web browsers are used to interpret special hypertext pages consisting of the HyperText Mark-up Language (HTML) and JavaScript so that they can be displayed in the given format. Some of the widely used web browsers are Internet Explorer, Netscape Navigator, Mozilla Firefox, Google Chrome, Lynx, Opera, Apple's Safari, etc.

5.6.1 Components of a Web Browser (Browser Architecture)

A Web browser comprises of three parts. These are controllers, client programs and interpreters.

- **Controller:** The controller obtains input from the keyboard or the mouse to access web pages with the help of a client program. After accessing the web pages, the controller uses one of the interpreters to display the web pages on the host screen.
- **Client Programs:** These are used to establish TCP sessions with the web server or the proxy server. To accomplish this task, the client programs make use of HTTP, FTP, Gopher or Telnet.
- **Interpreters:** These are used to display the web pages on the web user's screen. The interpreters which are used to translate web pages on the client's screen are HTML, CGI and JAVA. Such interpreters depend on the type of document. The HTML, which is a markup language and which allows the browser to change the format of the web pages, is used for scripting web pages. The HTML also helps store instructions along with the text, so that any browser can read the instructions and format the text according to the respective host machine.

5.7 WEB PAGE

The WWW is a subset of the Internet and comprises of a huge collection of documents stored in computers across the world. The web encompasses special sites called websites along the Internet, that support web browsing. By clicking on the links that appear on the webpage, one can navigate from one place to another. Hence, webpage can be defined as a single hypertext document written in HyperText Markup Language (HTML) and described in HTML basics. A webpage normally incorporates the basic information and links to navigate in the websites to which it belongs. Documents in the World Wide Web are classified into three types, namely static, dynamic and active documents.

5.7.1 Static Webpage

These are fixed content documents which perpetually provide the same information in response to all download requests from all web users. Static documents are stored in a web server to be accessed by the web client. The web client, on requesting for a web page, gets a copy of the same. The contents of such files are not subject to modification on part of the web user as the web user does not have

right to alter them. However, the web pages can be modified in the server per say. Thus, the static web pages display the same information to all the web users and provide hypertext links to perform navigation through static documents. Their biggest advantage is that they are cache friendly. This enables the web pages to display one copy of the same web page to many people simultaneously. However, it becomes difficult to maintain web pages in case of large sites as they demand consistency and updation.

5.7.2 Dynamic Webpage

These web pages provide interactive web navigation and help modify the content like text, images, form fields, etc. on a web page, depending on different contexts or conditions. The dynamic web pages make use of two types of inter-activities, which are enlisted in the following:

- **Client Side Scripting -** It is used to modify interface behaviours within a specific web page. This modification is based on the mouse or keyboard actions and is conducted at specified time intervals. The dynamic behaviour takes place within the presentation. The presentation technologies like JavaScript or ActionScript for Dynamic HTML (DHTML) and Flash for media types of the presentation are used. The client side scripting also facilitates the use of remote scripting in which the DHTML page requests for additional information from the server. The content is generated on the web client's machine in which the web browser retrieves a page from the server and processes the code embedded in the web page, so that the contents of the retrieved page can be displayed to the web user. Sometimes, the web browsers do not support the language and the commands of the scripting language, in the client-side dynamic pages.
- **Server Side Scripting -** It is used to modify the requested web page source amongst pages to either adjust the sequence or reload the web pages delivered to the browser. Server responses are based on certain conditions like data in a posted HTML form, parameters in the URL, the type of browser being used, the passage of time or a database or server state. Server side scripting dynamic web pages are designed with the help of server-side languages like PHP, Perl, ASP, JSP, etc.

Both the techniques may be used simultaneously to develop the dynamic web pages. The advantages of dynamic web pages are that these facilitate easy update of the web pages and faster web page loading. In the dynamic web pages, the content and the design are located separately, thereby allowing frequent modifications to the web pages including the text and image updates.

5.7.3 Active Documents

The programs that run at the client side are known as the active documents. Whenever a web client requests for an active document, the web server provides a copy of the same in the form of byte code. The document is now ready to be run

at the web client machine. As the active document is served in the binary form, compression and decompression can be applied at the server and the client side to reduce the bandwidth requirement and throughput.

NOTES

5.8 WORLD WIDE WEB

You have already studied at length about the World Wide Web (WWW) which is considered a major reason behind the popularity of Internet. WWW refers to a system of information and communications through which the users can access hypermedia information on servers. It is treasure trove of boundless information, in which all items have a reference through which they can be retrieved. In other words, any kind of information on any topic is readily available on the Internet. WWW comprises of a collection of websites which are publicly accessible. A website usually contains multiple pages which are replete with different types of information about different topics. Following are the constituents of a website:

- **Home Page:** This page tells the visitors across net about what is that. It also tells the visitors across net about what is that. The home page also locates the relevant sites on the net. It also provides detailed information about its service domain.



Fig. 5.10 *Home Page Screen*

The Figure 5.10 shows the Home Page of the site that provides various buttons to search the types. The website depicted here is meant for online searching of thesaurus, encyclopedia, style guide, word games and Spanish-English dictionary.

The web service provider helps establish a home page on the net. It is also known as electronic description of the organizations and its products and services. In real life, it is similar to a brochure or a catalog. An attractive home page grabs the attention of the visitors and that is why its cost per page is higher as compared to regular web pages. A home page can be setup with the help of Integrated Services Digital Network (ISDN) line as it generates HTML which represents the graphical interface.

For example, if you type www.google.com, it brings you the Google home page which is illustrated in Figure 5.11.

NOTES

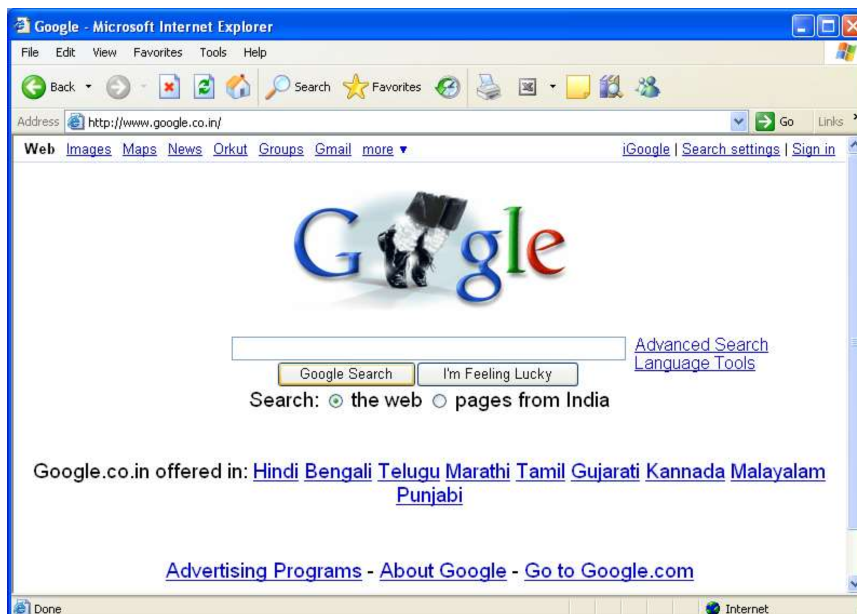


Fig. 5.11 *Google Home Page*

- **About Page:** This page provides the visitors with a detailed information of the launched website. For example, an online air ticket booking system gives a detailed description of its services for an air reservation domain.
- **Press Page:** This page on the site helps people interact with media. It also encourages the visitors to include the address, e-mail and phone/mobile numbers. It also publishes news on its website on a daily basis. Before launching any site, it is essential to check whether the press page, contact page and other attractive features are available on the website.
- **Contact Page:** The contact page provides the visitors with the city map, email address, phone numbers and other relevant information, so that the visitors can personally contact the organization or the service providers.

5.9 HOME PAGE

Home page is known as the first page of the web page. It is replete with a myriad of hyperlinks on its page. Creation of a home page connotes creating and launching of the website. This is a consequential task which is accomplished by arranging the website hosting, designing and coding of the website, monitoring the functioning of the site and by scrutinizing the website traffic. Creating the website takes into consideration, various factors which are to be implemented on the page. Launching the website is an important operation. This requires information pertaining to name, phone, URL description as well as the domain details. It should be ensured that the website must be kept in the right direction. A comprehensive user’s guide that conveys the relevant information of the website, must be provided to the user. This can be done after successfully launching the website. The more accurate the details

NOTES

are, the better the results would be. The task of launching the website can be carried out in the local listing of Google and Yahoo. This optimizes the search engine facilities for your website, which offers moderate list of options, searchable information and the third party data providers, such as SuperPages, YellowPages, CitySearches, etc. These search engines provide a great facility to recite the name of your website. Such search engines also offer a free Jumpstart program in which you can enter you website domain area, avail the 'WAY' (Who You Are) facility, get reviews and list hours etc. There are many factors that determine the success of a website on the net. The following factors should be taken into consideration while creating and launching a Website:

- **Message Board:** It is a type of forum through which visitors of the website interact with the site to enhance its popularity.
- **Search Engine:** This is a valuable retention tool which helps visitors search for relevant information, provided the site is enlisted with a famous search engine.
- **Polls:** This option on the website enables the visitors to vote as per their satisfaction. For instance, the users can assess the performance of the web services by giving their feedback with the help of the feedback option.
- **Guestbooks:** This website facility helps the users contact the organization for which the particular site is created. Through this facility, the website visitors can enter their name, email, comments and suggestions. Once this information reaches the organization, the respective executive of the organization contacts the visitors.
- **Data Entry Forms:** Through this option, the website visitors can place orders and can also provide request information. Date entry forms also facilitate storing of customer service data, which is later entered into a computerized database or spreadsheet by the organization.

Customer Record 1

Customer Information	
Customer No.	
First Name	
Last Name	
Sex	
Address Line 1	
Address Line 2	
City	
State	
Post Code	
Country	
	Submit to database or Add another customer record?

Fig. 5.12 Data Entry Form for Customer Record

NOTES

Figure 5.12 depicts the data entry form for customer record that provides various text fields, submit buttons, links to navigate to another page, combo boxes etc. If you click on the link Add another customer record? as specified in the figure, it will provide another set of customer record fields to fill the detail of customer information.

5.9.1 Creation of a Home Page

Creation of a home page requires eight steps which are as follows:

1. **Select and Register a Web Page Domain Name**

   First select a suitable website domain name to monitor the conflict issues of the same. Once a domain name is allotted to an organization or an individual, it can not be further allotted to anyone. The registration of domain name is unique and is carried out by Internet Corporation for Assigned Names and Numbers (ICANN), which is an -accredited domain name registrar, such as abc.com, xyz.com, etc. The free website hosting service is also available that can be availed without registering a domain name. The search engine does not provide its services if any website lacks its registered domain name.

2. **Select and Configure a Website Hosting Service**

   The hosting cost of website ranges from $100 to $250 every year. The cost varies from one website to another, depending on the websites' features such as ecommerce facilities, special processing requirements, high traffic volume options, etc. At this stage, web hosting is checked for control over content, security and usage of the site.

3. **Design, Code and Test the Website**

   A static website comprises of a single web page. It must have 'index.html' or 'index.htm'. A bare-bone format of the web page, which constitutes the HTML code, is as follows:

```
<HTML>
  <BODY> Hello Web! </BODY>
</HTML>
```

   There are various types of software tools, such as Adobe PhotoShop, Microsoft FrontPage, etc.

4. **Deploy the Website to the Host**

   At this stage, the file transfer program is uploaded in order to download the website. It also updates the pages between system units and website host computer.

NOTES

5. Security and Authentication

Before launching a website on the net, it is essential to implement the security and full-proof authentication of the site. The following security methods are required for the same:

- Login pages should be encrypted.
- Data validation should be conducted in the server-side.
- Managing the website with the help of encrypted connections.
- Website must be connected from secured network.
- Login credentials must not be shared.
- Maintaining a Password and key authentication.
- Use redundancy to protect the website.

After accomplishing the authentication only SSL is used with **http://URL** scheme is used. The login form POST is encrypted after every login process. The JavaScript is used for validating the web forms. The server side validation safeguards from malicious security crackers but in a limited way. The website must be equipped with encrypted connections because non-encrypted connections make the website susceptible to login or password sniffing or even man-in-middle-attacks.



Fig. 5.13 *Login Screen*

A secured connection must be associated with website and also secured proxy server must be used, such as **PuTTY** secure proxy or **OpenSSHproxy**. In order to maintain secure workstation, integrity auditing must be conducted. The server failover and backups should also be deployed as they diminish the possibility of server crashes. Data backup is also important in so far that it safeguards from loosing the client's data.

6. Online Payment Mode

Before launching any website, it is essential to set up an online payment mode, which is provided to the website. A website must be equipped with facilities like payment through credit cards or by a third-party such as Paypal, etc.

Other factors that should be considered while launching a website are following:

- A website is launched with the help of File Transfer Protocol Program (FTP). It is an economical option. The owner of website must instruct the web designer and the system analysts to implement FTP before the launch of the website.
- The web host firm place on the server must be provided on the site. For example, site owner provides the disk to the web host firm so that they can set up fee for the site.
- It is incumbent upon the owner of the website to remove the 'teething' problem before launching the website. Incoherent or incomplete website can discourage the visitors. For example, if an organization provides e-commerce services, it must ensure that up to date and relevant information is available on its site. The teething problem may lead to problem of set up and layout of the web screen.

A websites is written in HTML and is a collection of linked web pages on a web server which can be electronically accessed. Web server is a machine in which a website is located or hosted. It may be organization owned or Internet Service Provider (ISP) owned. The web pages may be owned by a university, a private company or an individual and are accessible to all people. Most of the websites have their own homepages that facilitate navigation by providing links to explore the details stored therein. The pages of a good website are organized using a common theme.

7. Launching the Website

Launch of the site is carried out after designing and completion of the site. It is essential to finalize the layout and style of the site before launching. It is significant to note that before the launch of the website, its domain should be registered.

8. Promote the Website

The information is sent on the web through search engines and their related directories. The promotion scheme must be published on the website at regular intervals. Therefore, this factor must be considered during creation of the website. The optimal way to promote the website is to update the visitors on the specific website with the pertinent information. For example, in case of online air ticket booking systems, any promotional scheme such as shifting the seat arrangement from economic to business class or changing the flight schedules etc. must be updated online to intimate the travellers and the visitors about the same.

It provides a point of entry to a Website with help. It also contains all relevant links of a particular website, so as to enable the user to explore the website for information available therein.

5.10 INTERNET AND URL

NOTES

URL denotes Uniform Resource Locator. It is the address of a document on the World Wide Web. Web browsers enable a person to enter either a known address in the web server or a specific document within that server. Addresses usually begin with http://, ftp://, gopher://, WAIS://, file:// etc. It is not feasible to maintain WWW without using the URLs. These are also used to represent hypermedia links and links to network services within the HTML documents. Any file or service on the Internet can be presented with the help of the URL. The first part of the URL that comes before the two slashes specifies the method of access or protocol being followed for communications between the browser and the web server. The second part coming after two slashes represents the address of the host machine, whose data or services are being sought. The remaining parts signify the names of the files, the port to connect to or the text to search for in a database. All the parts of an address for obtaining a file or service from a host machine in a URL are shown as a single unbroken line with no spaces and the locations of the host machines or websites that run www servers are typically named with a www at the beginning of the network address. The web browsers enable the users to access web services by specifying a URL and connecting to that document or service. Once the user gets connected with the web server, the web browsers select the hypertext in an HTML document and send a request to open a URL. Thus, hyperlinks are used not only to provide other texts and media in the same document but also to facilitate other network services. Web browsers are not simply web clients. They are full-featured FTP, Gopher and telnet clients.

5.11 SURFING THE INTERNET

The advent of Internet and WWW drastically changed the life of people for better. It has altogether altered the way people access information. With the help of WWW, a myriad of websites dealing with diverse subjects have come into existence. Millions of computers with billions of web pages are hooked to the Internet and are ready to provide information and knowledge pertaining to any subject or topic of your choice. To extract information from the Internet, we need to explore the Internet. Surfing the Internet entails searching for two types of materials, namely the textual and the non-textual materials. HyperText Transfer Protocol (HTTP) facilitates exchange of documents between two or more computers that are connected to the Internet. It is the web browser which has the ability to provide requisite documents in the form of web pages. As an Internet is a repertoire of information, one needs to be adept in digging out the relevant information from such a vast ocean of web pages. Basically, there are two popular methods for surfing the Internet. In the first method, you know the web address of the particular website in which the pertinent information can be searched for. In this case, you

NOTES

just need to key the web address in the form of URL in the address bar of your web browser. It enables you to see the homepage of the desired website which allows navigation of different web pages contained in that website. The second method necessitates the use of search engines, that is, the software systems which enable the users to search for information on the WWW using specific keywords. Internet users key some keywords in the space provided in the search engine page. Obtaining the desired information is contingent upon the keywords entered to a great extent. Surfing the Internet with the help of search engines has become an integral part of our life. Surfing the Internet also facilitates access to chat rooms where online discussions or chatting take place. However, surfing the Internet has several disadvantages too. Some notorious people misuse the Internet and hack into other peoples' private accounts. Widespread injection of Spam is also annoying.

5.12 SEARCH ENGINE

Search engines are the softwares that enable searching of the content available on Internet. A search engine is an information retrieval system which is used to access and retrieve information stored in WWW or a computer system attached to the Internet. Search engines also help minimize the time required to find the relevant information on the computer system. The computer system could be a standalone system or it could also be attached to the Internet. The search engines are popular amongst people as web search engines help explore information on the World Wide Web.

Search engines are the interface to a group of contents, which allow the users to type in the keywords, so that the engine can find several matching contents to the corresponding keywords out of millions of web pages. The keywords provided by the user are referred to as a search query. Several styles of search query syntax are used by the net users. Search query differs for different types of search engines, that is, some search engines enable users to enter two or three words separated by space, whereas others may require users to provide entire documents, pictures, sounds, and various forms of languages. Some search engines attempt to enhance the search queries to provide a quality set of items through a process known as query expansion.

`http://en.wikipedia.org/wiki/Image:Search-engine-diagram-en.svg`

5.12.1 Index-Based Search Engine

In case of such engines, the list of items to meet the criteria specified by the query, is typically sorted or indexed. Indexing the contents by relevance, that is, from the highest to lowest, minimizes the time needed to explore the desired information. Some search engines use probabilistic approach to rank the contents, depending on measures of similarity, popularity or authority. Boolean search engines typically provide contents which match exactly irrespective of the order in which the

NOTES

keywords are typed. However, the term boolean search engine may also allude to the use of boolean-style syntax. Thus, in order to provide a set of matching contents that are based on some criteria, the search engine will collect metadata concerning the group of contents under consideration, through a process called indexing. The advantage of indexing is that it calls for a smaller amount of computer storage.

5.12.2 Types and Characteristics

Some of the popular search engines with their types and characteristics are following:

- **Alta Vista** – It is a crawler type of search engine which comes up with results based on how many times the search words appear in the text. It searches the complete text.
- **Excite** – It is also a crawler type of search engine and it makes use of meta tags.
- **Google** – It is the most widely used search engine. It is also a crawler type and its results are based on the number of times, the other sites link to the ranked site.
- **Yahoo** – This is also a crawler type search engine and its functions are similar to that of Google.

5.13 UPLOADING AND DOWNLOADING

Uploading refers to sending of data from a local system to a remote system. like a server to keep a copy of the data there for various purposes. Downloading refers to retrieving of data from a remote system to a local system. Examples of remote systems include the web server, FTP server, email server or other similar systems. Information needs to be digitized for uploading or downloading.

In the case of uploading, the files are generally copied from a smaller peripheral system to a larger central system. For example, a mobile phone file can be easily uploaded to a personal computer. Similarly, files from personal computer can be uploaded again to a server. Small files take only a few seconds in uploading, whereas larger graphic files can take hours in uploading.

Downloading entails transfer of data from a central system to a smaller system. It is quite palpable that larger files take more time as compared to smaller files in downloading. Email is an interesting example of downloading and uploading in which emails in Inbox are downloaded from a server, whereas the replies are uploaded, so that they may be transmitted to the recipient. File Transfer Protocol (FTP) program is used to upload files to servers as well as to download files from remote locations. A number of programs are available to assist the users with uploading and downloading.

Check Your Progress

13. What can be incorporated in a Web page?
14. What do you mean by active document?
15. How a Home page can be setup?
16. Correlate the Home page with the Web page.
17. Describe the security methods required for authentication.
18. What is the significance of URL?
19. Why Internet surfing is required?
20. Why search engines are considered important while surfing Internet?
21. What is uploading and downloading?

5.14 TELNET

One of the most basic requirements of networked computers is the ability to facilitate a communication between a client terminal and recipient terminal on the remote systems, i.e., to provide communication to both the ends. That is, a user wants to invoke an application on a remote machine. There are a number of application protocols that allow this remote execution capability, most notably, the TELNET protocol. TELNET is a general-purpose client-server based application program that enables the connection to be established to a remote system in such a way that the local terminal appears to be a terminal at the remote system. In TELNET protocol, there is a standardized interface for interaction between a TELNET client and the TELNET server. Through this interface program, a TELNET client can access resources on the TELNET server as though the client were a local terminal connected to the server. It is one of the most popular ways to remotely control web servers on the Internet.

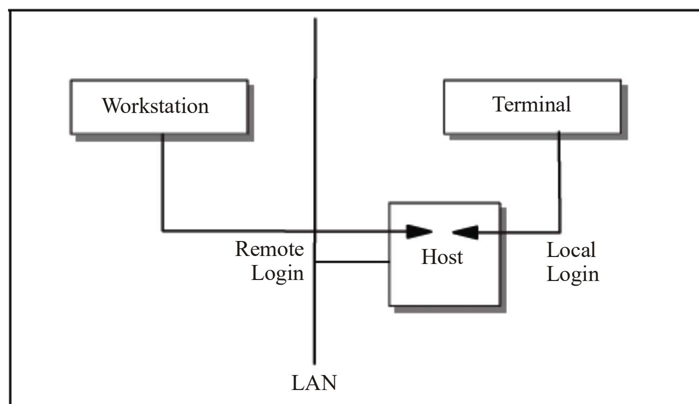


Fig. 5.14 *TELNET Operation*

NOTES

TELNET Operation

TELNET protocol is based on three basic concepts (see Figure 5.14):

- ***The Network Virtual Terminal (NVT) Concept*:** An NVT is an imaginary device with a basic structure common to a wide range of real terminals. Each host maps its own terminal characteristics to those of an NVT and assumes that every other host will do the same. It is simultaneously used by a horde of hosts to facilitate access to other servers.
- A ***Symmetric View*** of terminals and processes.
- ***Negotiation of Terminal Options*:** To provide services, TELNET hosts follow the principle of negotiated options though there are various options that can be negotiated. Sometimes, many hosts want to provide extra services beyond those available with NVT. Operational characteristics of their TELNET are instituted between the server and client by using a generic set of conventions through the 'DO, DONT, WILL, WONT' mechanism (see Figure 5.15).

  The process starts with two TELNET hosts verifying their mutual understanding. This is the initial negotiation where the hosts achieve minimum understanding. After this, they are capable of working on the minimum level implemented by the NVT. Once this initial negotiation is complete, the hosts negotiate additional options to provide capabilities beyond the minimum level implemented by NVT. Due to the symmetric model used by TELNET, both the host and the client can recommend additional options to be used.

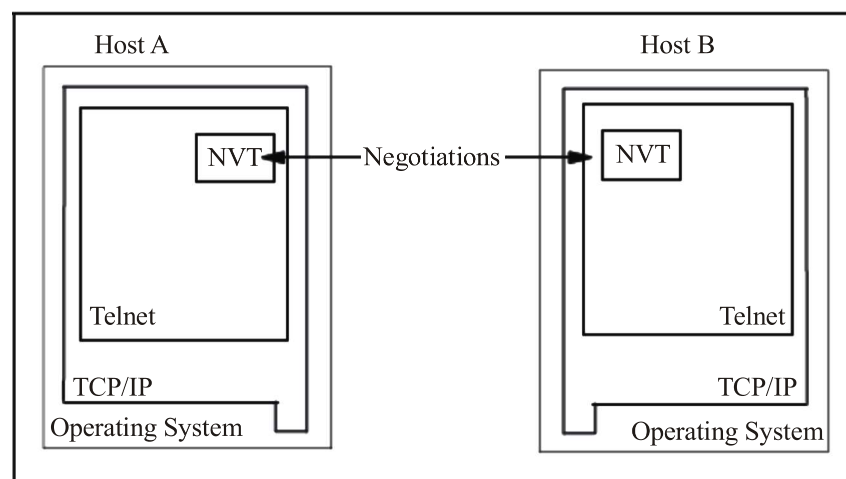


Fig. 5.15 *TELNET Negotiations*

Network Virtual Terminal

The Network Virtual Terminal (NVT) is an imaginary device consisting of a printer (or display) and a keyboard. Note that here the two peripherals play a contradictory role. The keyboard acts as an output device for producing outbound data to be

NOTES

sent over the TELNET connection. Quite contrary to this, the printer acts as an input device receiving the incoming data (see Figure 5.16).
The basic characteristics of an NVT are:

- Data is represented in 7-bit ASCII transmitted in 8-bit bytes.
- The NVT is a half-duplex device operating in a line-buffered mode.
- The NVT provides a local echo function.

  All of these can be negotiated by the two hosts and can be modified by mutually agreed options. For example, a local echo is favoured because of the lower network load and superior functioning.

  The TELNET client translates characters received from the local terminal into the NVT form and delivers them to the network. This is done using a universal interface called the NVT character set. The TELNET server then translates data and commands from the NVT form into the form acceptable by the remote computer.

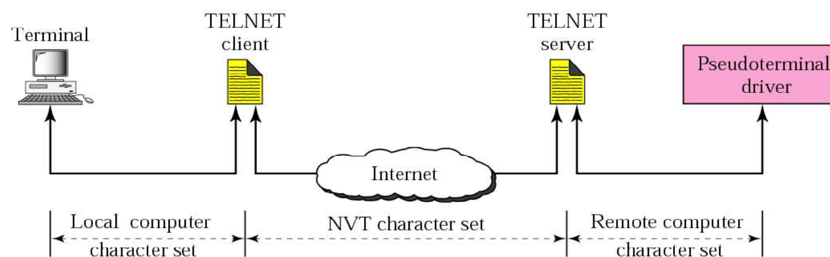


Fig. 5.16 *NVT*

NVT Character Set

There are two sets of characters used by NVT: Data and Control characters. Both are 8-bit bytes.

1. The format of data characters is shown in Figure 5.17.



Fig. 5.17 *Data Character*

2. The format of control characters is shown in Figure 5.18.

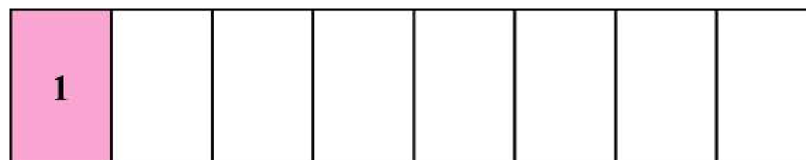


Fig. 5.18 *Control Character*

NOTES

3. Some NVT control characters.

   The following are some NVT control characters.

Character	Decimal	Binary	Meaning
EOF	236	11101100	End of file
EOR	239	11101111	End of record
SE	240	11110000	Suboption end
NOP	241	11110001	No operation
DM	242	11110010	Data mark
BRK	243	11110011	Break
IP	244	11110100	Interrupt process
AO	245	11110101	Abort output
AYT	246	11110110	Are you there?
EC	247	11110111	Erase character
EL	248	11111000	Erase line
GA	249	11111001	Go ahead
SB	250	11111010	Suboption begin
WILL	251	11111011	Agreement to enable option
WONT	252	11111100	Refusal to enable option
DO	253	11111101	Approval to option request
DONT	254	11111110	Denial of option request
IAC	255	11111111	Interpret (the next character) as control

Embedding

In TELNET, the data and control characters are transmitted through the same channel. This is accomplished by embedding the control characters in the data stream (see Figure 5.19).

A special control character called *Interpret As Control* (IAC) precedes each sequence of control characters. This is done to distinguish data from control characters.

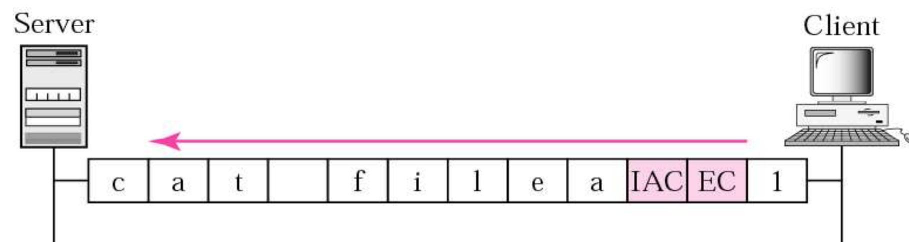


Fig. 5.19 *Embedding*

NOTES

Options

In TELNET, the client and server can negotiate options before or during the use of the service. Options are like extra features available to the user with a more sophisticated terminal. The following are the options available in TELNET.

Code	*Option*	*Meaning*
0	Binary	Interpret as 8-bit binary transmission
1	Echo	Echo the data received on one side to the other
3	Suppress go ahead	Suppress go-ahead signals after data
5	Status	Request the status of TELNET
6	Timing mark	Define the timing marks
24	Terminal type	Set the terminal type
32	Terminal speed	Set the terminal speed
34	Line mode	Change to line mode

Before any option can be used, option negotiation is required to be made between the client and the server on the networking protocol.

The following is the NVT character set for option negotiation.

Character	*Decimal*	*Binary*	*Meaning*
WILL	251	11111011	1. Offering to enable 2. Accepting a request to enable
WONT	252	11111100	1. Rejecting a request to enable 2. Offering to disable 3. Accepting a request to disable
DO	253	11111101	1. Approving an offer to enable 2. Requesting to enable
DONT	254	11111110	1. Disapproving an offer to enable 2. Approving an offer to disable 3. Requesting to disable

1. Offer to enable an option (see Figure 5.20).

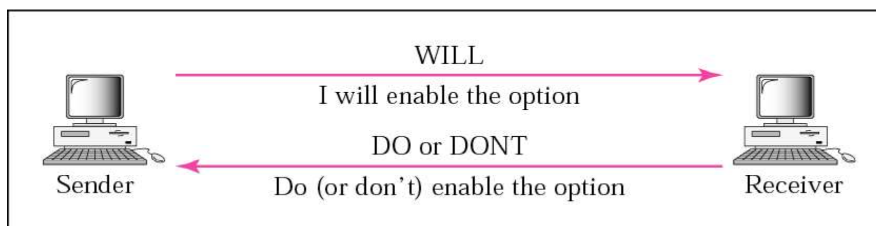


Fig. 5.20 *Offer to Enable an Option*

NOTES

2. Request to enable an option (see Figure 5.21).

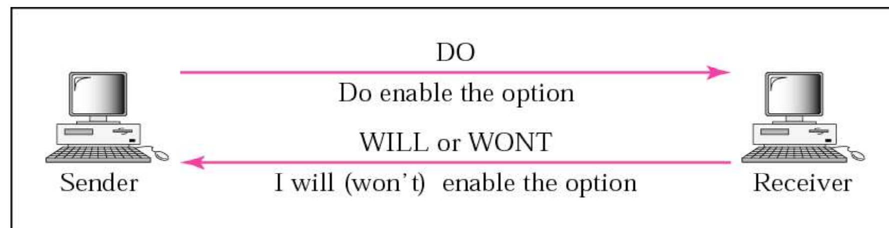


Fig. 5.21 Request to Enable an Option

3. Offer to disable an option (see Figure 5.22).

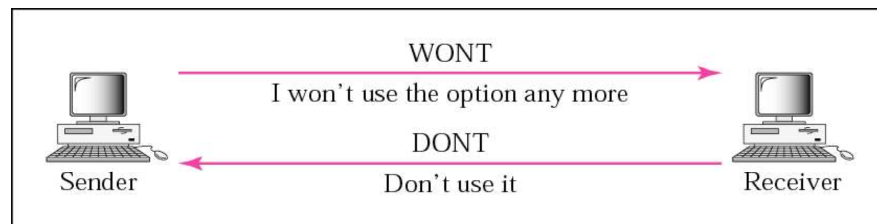


Fig. 5.22 Offer to Disable an Option

4. Request to disable an option (see Figure 5.23).

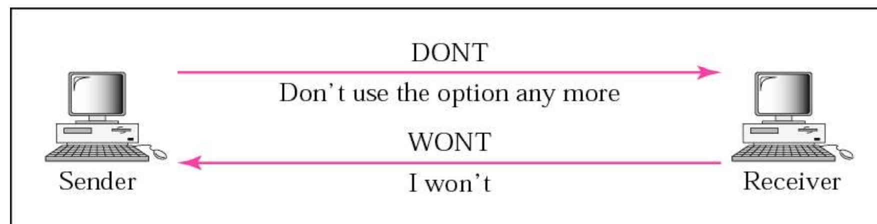


Fig. 5.23 Request to Disable an Option

Mode of Operation

TELNET provides the capability of running servers remotely and promotes remote network operations as well. This is where TELNET has an edge over the protocols. There are three modes in which TELNET can operate:

- **Default Mode**

  In case no other modes are invoked through option negotiation, the default mode is used. In this mode, the echoing is executed by the client. The user types a character; the client echoes the character on the screen but sends it only when the whole line is completed.

- **Character Mode**

  In the character mode, the user types a character; the client does not echo it but sends it to the server. The server echoes the character back to be

displayed on the client screen. In this mode, there can be a delay in echoing a character if the transmission time is too long. It also creates traffic for the network because three TCP segments must be sent for each character of data.

- **Line Mode**

  To gain mastery over the basic lacunae of the default and character mode, a new mode called the line mode has been proposed. In this mode, the client does the line editing and then transmits the entire line to the server.

Telnet provides the user with a unique login facility which enables the user to login remotely through remote computer to the host server. Telnet uses TCP protocol and hence it is connection-oriented. Through Telnet, the client is communicated to the destination port 23 to the server side. The protocol is specified to RFC 854. The RFC 854 supports telnet protocol specifications. Telnet is a virtual terminal facility which enables the users to connect with a remote system. Virtual terminals emulate a wide range of terminals, which reside in large or complex networks. It also provides a fundamental protocol to facilitate negotiation and it also controls various types of signalling and terminal protocols. There are four specific ways to access Telnet.. The terminal is connected to the TCP/IP terminal server with standard communication interface, for example, v-24/v-28 or EIA 232. These terminal types are registered with Internet Assigned Networks Authority (IANA). The Telnet is concerned with a number of parameters, such as speed, parity, flow control and types of terminals. Telnet is a packet switched network that came into existence as Telnet Inc. in the year 1947. The Telnet is used to access remote computer networks. It is an Internet service which enables the user of one host to log on to a remote host and interact as if the user has logged in locally. In order to accomplish this, it is indispensable to have the machine name and a valid user name. In order to access Telnet, certain commands are required, which are enlisted in the following:

- **Telnet Hostname:** This command opens a connection to the host name. For example, 'telnet abc.maths.edu' connects you to the machine named as telnet abc.maths.edu.
- **Telnet Address:** It initiates a connection to the host at IP address.
- **Echo:** This command echoes the detailed information which is to be sent to the port.

```
> telnet localhost echo
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
Hi echo service
Hi echo service
```

```
please repeat this line
please repeat this line
^]
```

NOTES

```
telnet> q
```

Connection closed.

- **Daytime:** The daytime command service returns the current time and Internet using telnet.

```
> telnet localhost daytime
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
Sun Feb 14 00:10:51 1999
```

Connection closed by foreign host.

- **Chargen**: This command is considered a character generator. It is an Internet protocol defined in RFC 864. It is used for testing and measurement purposes. The following command is issued for the same:

```
> telnet localhost chargen
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
```

!”#$%&’()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_‘abcdefgh

“#$%&’()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_`abcdefghi

#$%&’()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_‘abcdefghij

$%&’()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_‘abcdefghijk

%&’()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_‘abcdefghijkl

&’()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUV WXYZ[\]^_‘abcdefghijklm

‘()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_`abcdefghijklmn

()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU VWXYZ[\]^_‘abcdefghijklmno

)*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTU

VWXYZ[\]^_‘abcdefghijklmnop

* + , - . /
0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]
^_‘abcdefghijklmnopq

+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]
^_‘abcdefghijklmnopqr

,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]
^_‘abcdefghijklmnopqrs

-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]
^_‘abcdefghijklmnopqrst

[...]

- **Telnet Talk:** The DARPA telnet is an interactive communication protocol. It allows the remote session on the machine daemon service. It provides this service by typing the /usr/sbin/in.telnetd. It prompts you to enter the user name and the password. The following command is issued for telnet talk:

```
> telnet localhost
Trying 127.0.0.1...
Connected to localhost.
Escape character is '^]'.
Welcome to SuSE Linux 6.0 (i386) - Kernel 2.2.1 (ttyp5).

Tosh login: bb
Password:
Have a lot of fun...
Last login: Sat Feb 13 19:19:21 on ttyp0 from
Netwinder.suse.com.
You have mail.
Directory: /home/bb
Sun Feb 14 00:31:27 PST 1999
[tcshrc]
bb has logged on ttyp0 from netwinder.
bb has logged on ttyp5 from localhost.
[bb@Tosh:1] ~ >
```

- **Telnet mail.isp.com 110:** This Telnet command is used to send the email services through mail server to which POP3 server responds. The mail.isp.com is the Domain Name System of ISP and 110 is the mail server.

NOTES

Check Your Progress

22. Explain the basic concepts for TELNET protocol.
23. What is a network virtual terminal?
24. Describe the various modes in which TELNET can operate?
25. What is chargen?

5.15 USENET

Usenet is a worldwide distributed discussion system available on computers. It was developed from the general-purpose Unix-to-Unix Copy (UUCP) dial-up network architecture. Tom Truscott and Jim Ellis first thought on this idea in 1979, and it was finally established in 1980 at the University of North Carolina at Chapel Hill and Duke University, over a decade before the World Wide Web (WWW) went online, i.e., before the common users were given the access to the Internet, making it one of the oldest computer network communications systems which is still extensively used. Users read and post messages (called articles or posts, and collectively termed news) to one or more categories, known as newsgroups.

The name 'Usenet' is derived from the term 'Users Network'. The first Usenet group was named as NET.general, which immediately became net.general. The first commercial spam on Usenet was from immigration attorneys Canter and Siegel advertising green card services.

Usenet resembles a Bulletin Board System (BBS) in many respects and is the precursor to Internet forums that became widely used. Discussions are threaded, as with Web Forums and BBSs, though posts are stored on the server sequentially. The key difference between a BBS or Web Forum and Usenet is the absence of a central server and dedicated administrator. Usenet is distributed among a large, constantly changing conglomeration or assembly of servers that store and forward messages to one another by means of 'News Feeds'. Individual users may read messages from and post messages to a local server, which may be operated by anyone.

Usenet is culturally and historically significant in the networked world, having given rise to, or popularized, many widely recognized concepts and terms, such as 'FAQ', 'Flame', 'Sockpuppet', and 'Spam'. In the early 1990s, shortly before access to the Internet became commonly affordable, Usenet connections through FidoNet dial-up BBS networks made long-distance or worldwide discussions and other communication widespread, not requiring or demanding a server, but just local telephone service.

NOTES

The following Figure 5.24 illustrates the simple diagram or structure of Usenet servers and clients. The dots on the servers represent the groups they carry. Arrows between servers indicate newsgroup group exchanges (feeds). Arrows between clients and servers indicate that a user is subscribed to a certain group and reads or submits articles.

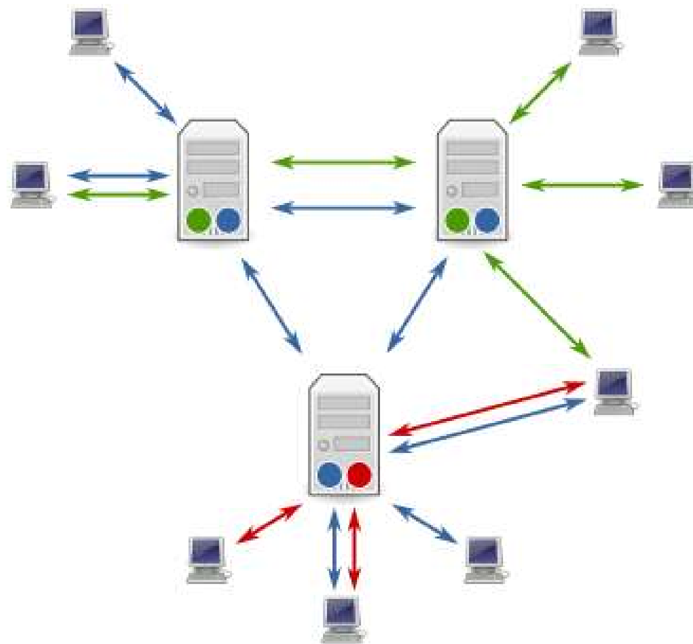


Fig. 5.24 *Simple Diagram or Structure of Usenet Servers and Clients*

The format and transmission of Usenet articles is similar to that of Internet e-mail messages. The difference between the two is that Usenet articles can be read by any user whose news server carries the group to which the message was posted, as opposed to e-mail messages, which have one or more specific recipients. Nowadays, the Usenet is not much used in comparison to Internet Forums, Blogs, Mailing Lists and Social Media.

Usenet is, therefore, a set of protocols for generating, storing and retrieving news 'Articles', which are similar to the Internet mail messages and for exchanging them among a readership which is potentially widely distributed. These protocols most commonly use a 'Flooding Algorithm' which propagates copies throughout a network of participating servers. Whenever a message reaches a server, that server forwards the message to all its network neighbours that have not yet seen the article. Only one copy of a message is stored per server, and each server makes it available on demand to the (typically local) readers able to access that server. The collection of Usenet servers has thus a definite peer-to-peer character in that they share resources by exchanging them.

RFC 850 was the first formal specification of the messages exchanged by Usenet servers. It was superseded by RFC 1036 and subsequently by RFC 5536 and RFC 5537.

On the Internet, Usenet is transported through the Network News Transfer Protocol (NNTP) on TCP Port 119 for standard, unprotected connections and on TCP port 563 for SSL encrypted connections.

NOTES

In addition, the Usenet platform permits its millions of users to save files, termed as 'Binaries' on the news servers for other members to download. Therefore, the Usenet platform allows the users to chat and share files, and also provides complete, unrestricted download and access speeds generally free exchange of content and information.

5.16 AOL

AOL, formalised as Aol., formerly a company known as AOL Inc. and originally known as America Online, is an American web portal and online service provider based in New York City. It is a brand marketed by Verizon Media.

The service traces its history to an online service known as PlayNET, which hosted multi-player games for the Commodore 64. PlayNET licensed their software to a new service, Quantum Link (Q-Link), who went online in November 1985. PlayNET shut down shortly thereafter. The initial Q-Link service was similar to the original PlayNET, but over time Q-Link added many new services. When a new IBM PC client was released, the company focused on the non-gaming services and launched it under the name America Online. The original Q-Link was shut down on November 1, 1995, while AOL grew to become the largest online service, displacing established players like CompuServe and The Source. By 1995, AOL had about three million active users.

AOL was one of the early pioneers of the Internet in the mid-1990s, and the most recognized brand on the web in the United States. It originally provided a dial-up service to millions of Americans, as well as providing a web portal, e-mail, instant messaging and later a web browser following its purchase of Netscape. In 2001, at the height of its popularity, it purchased the media conglomerate Time Warner in the largest merger in U.S. history. AOL rapidly declined thereafter, partly due to the decline of dial-up and rise of broadband. AOL was eventually spun off from Time Warner in 2009, with Tim Armstrong appointed the new CEO. Under his leadership, the company invested in media brands and advertising technologies.

On June 23, 2015, AOL was acquired by Verizon Communications.

AOL Mail

AOL Mail, formalised as AOL Mail, is a free web-based email service provided by AOL, a division of Verizon Communications. In 1993, both America Online (AOL) and Delphi started connecting their proprietary e-mail services to the Internet.

NOTES

Features of AOL Mail: The following are the features of AOL Mail:

- Email attachment limit: 25 MB
- Max mailbox size: 1,000 new messages, 4,000 old messages and 4,000 sent messages per screen name. Max 250 GB if all messages have 25 MB attachments.
- Supported protocols: POP3, SMTP, IMAP.
- Link to other e-mail accounts from other service providers (such as, Gmail and Hotmail).
- Ads are displayed while working with the e-mail account. Embedded links within emails are automatically disabled and can only be activated by the email user.
- Spam protection.
- Virus protection.
- Spell checking.
- E-mail unsent capability (for mails sent to another AOL or AIM mailbox).
- Domains: @aol.com and additionally @ygm.com (short for you've got mail), @games.com, and @wow.com.
- Supports SSL/HTTPS after login.

If an AOL Mail account is inactive for 90 days, it may become deactivated, at which point any e-mails sent to it may not be delivered and may be returned to sender. After 180 days of inactivity, the account may be deleted automatically.

Check Your Progress

26. Explain the features of Usenet.
27. On the Internet, how Usenet is transported?
28. What is AOL?

5.17 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Internetworking is a scheme for interconnecting multiple networks of dissimilar technologies. To interconnect multiple networks of dissimilar technologies use both additional hardware and software. This additional hardware is positioned between networks and software on each attached computer. This system of interconnected networks is called an *Internetwork* or an Internet.

NOTES

2. The word Internet is a short form of a complete word Internetwork or interconnected network. Therefore, it can be said that the Internet is not a single network, but a collection of networks. The commonality between them in order to communicate with each other is TCP/IP.
3. ARPANET was basically a network based on leased lines connected by special switching nodes, known as Internet Message Processors (IMP). ARPANET provided interconnection of various packet-switching nodes (PSN) located across continental USA and Western Europe using 56 Kbps leased lines. ARPANET provided connection to minicomputers running a protocol known as 1822 (after the number of a report describing it) and dedicated it to the packet-switching task.
4. The Domain Name System (DNS) is a distributed database that provides e-mail routing information. It is used by TCP/IP protocols. These protocols map between IP addresses and hostnames. The domain name extension type is tagged with distributed database because no single website across a network can access all the information. Each site runs a server program for sites and maintains its own database of information. Therefore, collecting information is possible through distributed method that creates a mechanism for clients and servers to communicate with each other.
5. E-mail is the prime Internet service that facilitates services to people or users across the world. Full Internet connectivity is not required for this. The header and body of the message make an e-mail message. The header contains the information where the message is to be sent and the complete path for reaching the destination, date and return path. The body of the message is the actual message that has to be sent. The syntax of an e-mail address is user@ subdomain.subdomain.domain, e.g., **abc@gmail.com**. A service provider must be connected with leased line, dial-up or connection with any network for sending e-mail.
6. The FTP operations can be performed by issuing FTP commands at the command prompt or by using FTP utility running under a graphical user interface on Windows OS.
7. Telnet is used to connect remote network computers. It is the Internet service that executes commands on remote host as if you are going to log in locally. For this, the machine name and valid user name are required to be connected.
8. The Gopher protocol supports client–server software that searches files on the Internet. A Gopher client is required for validating and testing of Gopher publishing service. Gopher browsers initially display the text-based files.
9. Finger service gives information about users, for example, username, person's first name and last name, information about recently logged in and also where they logged in. Finger is also used to get a list of users who are currently logged into the host.

10. The following are the primary advantages of e-mail:
    - It conducts paperless communication of messages quickly.
    - It ensures simultaneous transmission of messages to several users. The messages may comprise of pictures, video, film clips, text, animation or even a combination of them. Voice and audio messages can also be transmitted this way.
    - The email messages can also be printed, prioritized, forwarded and stored.
    - Public bulletin boards can be created in which every member of the organization can post and view messages. This can also be accomplished in the case of shared text messages and application files used widely across computer platforms.
    - It allows delivery and receiving of faxes and meetings can also be scheduled through email.
11. The G-mail interface provides the facility for opening of a new account, for which it provides a registration form to be filled up by the user. In accordance with the procedure, the user mentions his personal information, email id and password in the form. Thereafter, he gets registered and obtains an email address.
12. The Inbox folder contains all your previous email messages and also enables you to read the new ones. You also have an option of deleting the previous messages or transferring them to some other folders also.
13. A web page normally incorporates the basic information and links to navigate in the websites to which it belongs. Documents in the World Wide Web are classified into three types, namely static, dynamic and active documents.
14. The programs that run at the client side are known as the active documents. Whenever a web client requests for an active document, the web server provides a copy of the same in the form of byte code.
15. A Home page can be setup with the help of Integrated Services Digital Network (ISDN) line as it generates HTML which represents the graphical interface.
16. Home page is known as the first page of the web page. It is replete with a myriad of hyperlinks on its page.
17. The following security methods are required for the same:
    - Login pages should be encrypted.
    - Data validation should be conducted in the server-side.
    - Managing the website with the help of encrypted connections.
    - Website must be connected from secured network.

- Login credentials must not be shared.
- Maintaining a Password and key authentication.
- Use redundancy to protect the website.

NOTES

18. URL denotes Uniform Resource Locator. It is the address of a document on the World Wide Web. Web browsers enable a person to enter either a known address in the web server or a specific document within that server. Addresses usually begin with http://, ftp://, gopher://, WAIS://, file:// etc. It is not feasible to maintain WWW without using the URLs.
19. Surfing the Internet entails searching for two types of materials, namely the textual and the non-textual materials. HyperText Transfer Protocol (HTTP) facilitates exchange of documents between two or more computers that are connected to the Internet.
20. Search engines are the softwares that enable searching of the content available on Internet. A search engine is an information retrieval system which is used to access and retrieve information stored in WWW or a computer system attached to the Internet. Search engines also help minimize the time required to find the relevant information on the computer system.
21. Uploading refers to sending of data from a local system to a remote system. like a server to keep a copy of the data there for various purposes. Downloading refers to retrieving of data from a remote system to a local system.
22. TELNET protocol is based on three basic concepts.
    - The Network Virtual Terminal (NVT) concept: An NVT is an imaginary device with a basic structure common to a wide range of real terminals. Each host maps its own terminal characteristics to those of an NVT and assumes that every other host will do the same. It is simultaneously used by a horde of hosts to facilitate access to other servers.
    - A symmetric view of terminals and processes.
    - Negotiation of terminal options: To provide services, TELNET hosts follow the principle of negotiated options though there are various options that can be negotiated. Sometimes, many hosts want to provide extra services beyond those available with NVT. Operational characteristics of their TELNET are instituted between the server and client by using a generic set of conventions through the 'DO, DONT, WILL, WONT' mechanism.
23. The Network Virtual Terminal (NVT) is an imaginary device consisting of a printer (or display) and a keyboard. Note that here the two peripherals play a contradictory role. The keyboard acts as an output device for producing outbound data to be sent over the TELNET connection.

NOTES

24. There are three modes in which TELNET can operate:
    - Default Mode

      In case no other modes are invoked through option negotiation, the default mode is used. In this mode, the echoing is executed by the client. The user types a character; the client echoes the character on the screen but sends it only when the whole line is completed.
    - Character Mode

      In the character mode, the user types a character; the client does not echo it but sends it to the server. The server echoes the character back to be displayed on the client screen. In this mode, there can be a delay in echoing a character if the transmission time is too long. It also creates traffic for the network because three TCP segments must be sent for each character of data.
    - Line Mode

      To gain mastery over the basic lacunae of the default and character mode, a new mode called the line mode has been proposed. In this mode, the client does the line editing and then transmits the entire line to the server.
25. This command is considered a character generator. It is an Internet protocol defined in RFC 864. It is used for testing and measurement purposes.
26. Usenet is a worldwide distributed discussion system available on computers. It was developed from the general-purpose Unix-to-Unix Copy (UUCP) dial-up network architecture. Usenet resembles a Bulletin Board System (BBS) in many respects and is the precursor to Internet forums that became widely used. Usenet uses a set of protocols for generating, storing and retrieving news 'Articles', which are similar to the Internet mail messages and for exchanging them among a readership which is potentially widely distributed.
27. On the Internet, Usenet is transported through the Network News Transfer Protocol (NNTP) on TCP Port 119 for standard, unprotected connections and on TCP port 563 for SSL encrypted connections.
28. AOL, formalised as Aol., formerly a company known as AOL Inc. and originally known as America Online, is an American web portal and online service provider based in New York City. It is a brand marketed by Verizon Media.

5.18 SUMMARY

- The Internet, World Wide Web and Information Super Highway are terms which have the lives of millions of people all over the world. The widespread impact of Internet across the globe could not be possible without the

NOTES

development of Transmission Control Protocol/Internet Protocol (TCP/IP). This protocol suite is developed specifically for the Internet.

- During late 60s and 70s, organizations were inundated with many different LAN and WAN technologies such as packet switching technology, collision-detection local area networks, hierarchical enterprise networks, and many other excellent technologies. The major drawbacks of all these technologies were that they could not communicate with each other without expensive deployment of communications devices.
- Multiple networking models were available as a result of the research and development efforts made by many interest groups. This paved the way for development of another aspect of networking known as protocol layering. This allows applications to communicate with each other. A complete range of architectural models were proposed and implemented by various research teams and computer manufacturers.
- *Internetworking* is a scheme for interconnecting multiple networks of dissimilar technologies. To interconnect multiple networks of dissimilar technologies use both additional hardware and software. This additional hardware is positioned between networks and software on each attached computer. This system of interconnected networks is called an *Internetwork* or an *Internet*.
- ARPANET, a project of DARPA, introduced the world of networking with protocol suite concepts such as layering, well before ISO's initiative in this direction. DARPA continued its research for an Internetworking protocol suite. This may be seen in the early NCP (Network Control Program) host-to-host protocol to the TCP/IP protocol suite, which took its current form around 1978. DARPA was well known for its pioneering of packet switching over radio networks and satellite channels and ARPANET was declared an operational network with responsibility of administering it to Defense Communications Agency (DCA) in 1975. TCP/IP had not yet been developed.
- The first real implementation of the Internet was when DARPA converted the machines of its research network ARPANET to use the new TCP/IP protocols. After this transition which started in 1980 and finished in 1983, DARPA demanded that all computers willing to connect to its ARPANET must use TCP/IP. The US military adopted TCP/IP as standard protocol in 1983 and recommended that all networks connected to the ARPANET conform to the new standards.
- The word 'Internet' is a short form of a complete word Internetwork or interconnected network. Therefore, it can be said that the Internet is not a single network, but a collection of networks. The commonality between them in order to communicate with each other is TCP/IP.

- The Internet consists of the following groups of networks:
  - o **Backbones:** These are large networks that exist primarily to interconnect other networks. Some examples of backbones are NSFNET in the USA, EBONE in Europe and large commercial backbones.
  - o **Regional Networks:** These connect, for example, universities and colleges. ERNET (Education and Research Network) is an example in the Indian context.
  - o **Commercial Networks:** They provide access to the backbones to subscribers, and networks owned by commercial organizations for internal use and also have connections to the Internet. Mainly, Internet Service Providers come into this category.
  - o **Local Networks:** These are campus-wide university networks.
- The Domain Name System (DNS) is a distributed database that provides e-mail routing information. It is used by TCP/IP protocols. These protocols map between IP addresses and hostnames.
- The Domain Name Extension (DNE) is defined as the complete address of hosting services provided on the sites. In the beginning, the Internet configuration used numeric IP address, which was a very cumbersome task. To overcome this problem, symbolic host names came into existence. For instance, initially, it was typed as TELNET 10.12.7.14 but nowadays, TELNET MyHost command is issued. With this command, the mapping between machine names and IP addresses has also become centralized and coordinated.
- E-mail is the prime Internet service that facilitates services to people or users across the world. Full Internet connectivity is not required for this. For example, an electronic address provides these services to FTP sites through which mail can be exchanged. Other Internet services, such as IP address resolver, Archie lookup, WHOIS service is done via e-mail.
- FTP is also prime Internet service that acts as protocol and transfers files over TCP/IP network (Internet, UNIT, etc.). Once HTML page is developed on a local machine for a website, it is first uploaded to the Web server through FTP. Local machine is the machine on which you are initially logged into. It includes functions to log on to the network, gives a list of directories and copies files. FTP transfer is possible by entering URL preceded with ftp:// within address bar of a web browser. The FTP operations can be performed by issuing FTP commands at the command prompt or by using FTP utility running under a graphical user interface on Windows OS. FTP tasks can be performed through a browser. For example, type an IE address bar URL as ftp:// to get ftp services. For example, ftp://YourLoginName@IPaddress.

NOTES

- Telnet is used to connect remote network computers. It is the Internet service that executes commands on remote host as if you are going to log in locally. For this, the machine name and valid user name are required to be connected. The commands that are issued on telnet are as follows:

  ***Telnet Hostname*:** A connection to the host name is opened by this command. For example, issuing the command as 'telnet abc.maths.edu' with that machine which keeps the required information of abc.maths.edu site can connect you.

  ***Telnet Address*:** It gives the IP address of the connected host.

- Archie is a program that searches files anywhere on the Net by filename. This facility is maintained by a database with the Internet sites accessible via anonymous FTP.
- Gopher protocol supports client–server software that searches files on the Internet. A Gopher client is required for validating and testing of Gopher publishing service. For example, WS Gopher 1.2 is available on the Internet as shareware. The server based text files are hierarchically organized and viewed by end-users.
- Finger service gives information about users, for example, username, person's first name and last name, information about recently logged in and also where they logged in. But the users must enter the required information where they get registration for particular e-mail services. Finger is also used to get a list of users who are currently logged into the host.
- WWW provides hypertext access to documents located anywhere on the Internet. It is a very successful distributed information system. It is basically client–server data transfer protocol that communicates via application level protocol. Its structural components are clients–browsers, servers and caches. The Internet and semantic components include HyperText Transfer Protocol (HTTP), HyperText Markup Language (HTML) eXtensible Markup Language (XML) and Uniform Resource Identifiers (URIs). The clients who get various sites requested to the server via HTTP determine the structure of WWW. Then web pages constructs HTML consists of graphics and sound embedded files. For running the complete system, TCP/IP, DNS networking protocols are required.
- Electronic mail is one of the most popular network services. The use of e-mail is considered the foremost reason behind the popularity of Internet. The proliferation of cyber cafés can be attributed to e-mail or World Wide Web. E-mail provides an efficient and fast means of communication with relatives, friends or colleagues throughout the world.

NOTES

- Certain equipments which are required to access the Internet are as follows: Amongst these, some of them are mandatory and some are optional:
  - o **Computer:** A computer which is used to browse the Internet may either be a personal computer with Pentium processor or a Macintosh. It should have enough power and memory concomitant with multimedia features. Though 128 MB RAM is sufficient to have access to Internet but 512 MB RAM or more is recommended. Now-a-days, devices like smart phones, mobile phones, Pocket PCs, etc., are also used to browse the Internet.
  - o **MODEM:** It stands for MOdulator/DEModulator. This may either be internally built in or externally connected. The MODEM is a device that converts data in binary code used by the computer, to an analog signal that can be transmitted over the telephone network and vice versa.

    With the help of telephone lines, millions of computers worldwide are connected with one another, either directly or indirectly. In order to connect with the Internet Service Provider (ISP), these connections require the regular dial up telephone lines or dedicated higher capacity telephone lines like leased lines, ISDN lines, etc.
  - o **Internet Account with a Service Provider:** An account with a service provider is essential to create a link between the user's computer and the Internet. A service provider, which is popularly referred to as ISP (Internet Service Provider), signifies phone or cable companies that provide last mile connectivity. It may also refer to a cable line from the subscriber's home to his office and also to an exchange for long distance connectivity based on monthly or annual charges.
  - o **Widely used Current Standard Broadband Technologies:** These technologies are DSLs, that is, Digital Subscriber Line (DSL) and cable modems. However, recent technologies like VDSL and optical fiber connections are also gradually becoming popular in providing Internet access in a much more cost-effective way than copper wire technology. Wi-Fi networks are also used to provide Internet connections. However, these are not served in the areas by cable or ADSL. The WiMAX has been gaining popularity with regard to mobile and stationary broadband access.
  - o **Internet Browser Software:** It is the software tool which enables a user to browse the Internet with the help of web addresses or URLs. A few of the widely used browsers are Internet Explorer version 7 or 8 (IE), Netscape, Mozilla Firefox, Chrome, AOL, Opera, etc.
  - o **Anti-Virus Software:** These are used to protect the user from the onslaught of the nasty programs that obtain access to the user's terminal

NOTES

when he is surfing the network or downloading contents from there. Some examples of anti-virus software are Symantec, Norton, McAfee, etc.

- o **Email Software:** The email software may be chosen from the Outlook or Outlook Express. Google, Yahoo and Hotmail offer free web-mail for the same.
- o **Plug-In Software:** It is considered an add-on to the user terminal. It enables the user to avail services like music, video, multimedia, etc. on the Internet. The most popular plug-in-softwares include Real Audio music player, Macromedia Flash Player, Windows Media Player, Apple Quick Time, Java Virtual Machine, etc.
- o **Stereo Speakers, Microphone and Webcam:** These equipments enable the user to play sounds, videos, to conduct Internet telephoning and to send images to other users connected to the Internet.

- A browser is a software, which your computer uses to view WWW documents and access the Internet. The browser program residing in your computer provides you with the facilities like text formatting, hypertext links, images, sounds, motion and other features. Internet Explorer and Netscape are considered the most widely used browsers. Browsers have sub programs called plug-ins to handle the documents found on the Web. It may also have other plug-ins stored elsewhere in the computer.
- The WWW is a subset of the Internet and comprises of a huge collection of documents stored in computers across the world. The web encompasses special sites called websites along the Internet, that support web browsing. By clicking on the links that appear on the webpage, one can navigate from one place to another. Hence, webpage can be defined as a single hypertext document written in HyperText Markup Language (HTML) and described in HTML basics. A webpage normally incorporates the basic information and links to navigate in the websites to which it belongs. Documents in the World Wide Web are classified into three types, namely static, dynamic and active documents.
- Static Webpage these are fixed content documents which perpetually provide the same information in response to all download requests from all web users. Static documents are stored in a web server to be accessed by the web client. The web client, on requesting for a web page, gets a copy of the same. The contents of such files are not subject to modification on part of the web user as the web user does not have right to alter them.
- Dynamic Webpage these web pages provide interactive web navigation and help modify the content like text, images, form fields, etc. on a web page, depending on different contexts or conditions.

NOTES

- Active Documents the programs that run at the client side are known as the active documents. Whenever a web client requests for an active document, the web server provides a copy of the same in the form of byte code. The document is now ready to be run at the web client machine. As the active document is served in the binary form, compression and decompression can be applied at the server and the client side to reduce the bandwidth requirement and throughput.
- Home page is known as the first page of the web page. It is replete with a myriad of hyperlinks on its page. Creation of a home page connotes creating and launching of the website. This is a consequential task which is accomplished by arranging the website hosting, designing and coding of the website, monitoring the functioning of the site and by scrutinizing the website traffic. Creating the website takes into consideration, various factors which are to be implemented on the page. Launching the website is an important operation. This requires information pertaining to name, phone, URL description as well as the domain details. It should be ensured that the website must be kept in the right direction. A comprehensive user's guide that conveys the relevant information of the website, must be provided to the user. This can be done after successfully launching the website.
- The following factors should be taken into consideration while creating and launching a Website:
  - o **Message Board:** It is a type of forum through which visitors of the website interact with the site to enhance its popularity.
  - o **Search Engine:** This is a valuable retention tool which helps visitors search for relevant information, provided the site is enlisted with a famous search engine.
  - o **Polls:** This option on the website enables the visitors to vote as per their satisfaction. For instance, the users can assess the performance of the web services by giving their feedback with the help of the feedback option.
  - o **Guestbooks:** This website facility helps the users contact the organization for which the particular site is created. Through this facility, the website visitors can enter their name, email, comments and suggestions. Once this information reaches the organization, the respective executive of the organization contacts the visitors.
  - o **Data Entry Forms:** Through this option, the website visitors can place orders and can also provide request information. Date entry forms also facilitate storing of customer service data, which is later entered into a computerized database or spreadsheet by the organization.

NOTES

- URL denotes Uniform Resource Locator. It is the address of a document on the World Wide Web. Web browsers enable a person to enter either a known address in the web server or a specific document within that server. Addresses usually begin with http://, ftp://, gopher://, WAIS://, file:// etc. It is not feasible to maintain WWW without using the URLs. These are also used to represent hypermedia links and links to network services within the HTML documents. Any file or service on the Internet can be presented with the help of the URL. The first part of the URL that comes before the two slashes specifies the method of access or protocol being followed for communications between the browser and the web server. The second part coming after two slashes represents the address of the host machine, whose data or services are being sought. The remaining parts signify the names of the files, the port to connect to or the text to search for in a database.
- Search engines are the softwares that enable searching of the content available on Internet. A search engine is an information retrieval system which is used to access and retrieve information stored in WWW or a computer system attached to the Internet. Search engines also help minimize the time required to find the relevant information on the computer system.
- Uploading refers to sending of data from a local system to a remote system. like a server to keep a copy of the data there for various purposes. Downloading refers to retrieving of data from a remote system to a local system. Examples of remote systems include the web server, FTP server, email server or other similar systems. Information needs to be digitized for uploading or downloading.
- Downloading entails transfer of data from a central system to a smaller system. It is quite palpable that larger files take more time as compared to smaller files in downloading. Email is an interesting example of downloading and uploading in which emails in Inbox are downloaded from a server, whereas the replies are uploaded, so that they may be transmitted to the recipient. File Transfer Protocol (FTP) program is used to upload files to servers as well as to download files from remote locations. A number of programs are available to assist the users with uploading and downloading.
- TELNET is a general-purpose client-server based application program that enables the connection to be established to a remote system in such a way that the local terminal appears to be a terminal at the remote system. In TELNET protocol, there is a standardized interface for interaction between a TELNET client and the TELNET server. Through this interface program, a TELNET client can access resources on the TELNET server as though the client were a local terminal connected to the server. It is one of the most popular ways to remotely control web servers on the Internet.

- Usenet is a worldwide distributed discussion system available on computers. It was developed from the general-purpose Unix-to-Unix Copy (UUCP) dial-up network architecture. Tom Truscott and Jim Ellis first thought on this idea in 1979, and it was finally established in 1980 at the University of North Carolina at Chapel Hill and Duke University, over a decade before the World Wide Web (WWW) went online, i.e., before the common users were given the access to the Internet, making it one of the oldest computer network communications systems which is still extensively used. Users read and post messages (called articles or posts, and collectively termed news) to one or more categories, known as newsgroups.
- AOL, formalised as Aol., formerly a company known as AOL Inc. and originally known as America Online, is an American web portal and online service provider based in New York City. It is a brand marketed by Verizon Media.

  The service traces its history to an online service known as PlayNET, which hosted multi-player games for the Commodore 64. PlayNET licensed their software to a new service, Quantum Link (Q-Link), who went online in November 1985. PlayNET shut down shortly thereafter. The initial Q-Link service was similar to the original PlayNET, but over time Q-Link added many new services.

5.19 KEY WORDS

- **Telnet Hostname:** A connection to the host name is opened by this command. For example, issuing the command as 'telnet abc.maths.edu' with that machine which keeps the required information of abc.maths.edu site can connect you.

  Telnet Address: It gives the IP address of the connected host.
- **Archie:** Archie is a program that searches files anywhere on the Net by filename. This facility is maintained by a database with the Internet sites accessible via anonymous FTP.
- **Gopher:** Gopher protocol supports client–server software that searches files on the Internet. A Gopher client is required for validating and testing of Gopher publishing service. For example, WS Gopher 1.2 is available on the Internet as shareware. The server based text files are hierarchically organized and viewed by end-users.
- **Finger:** Finger service gives information about users, for example, username, person's first name and last name, information about recently logged in and also where they logged in.

NOTES

- **Email:** The term email connotes the basic communication facility provided by the Internet to its users to send and receive messages in any part of the world.
- **MODEM:** It stands for MOdulator/DEModulator. This may either be internally built in or externally connected. The MODEM is a device that converts data in binary code used by the computer, to an analog signal that can be transmitted over the telephone network and vice versa.
- **Contact Page:** The contact page provides the visitors with the city map, email address, phone numbers and other relevant information, so that the visitors can personally contact the organization or the service providers.
- **Search Engine:** This is a valuable retention tool which helps visitors search for relevant information, provided the site is enlisted with a famous search engine.
- **Network Virtual Terminal:** The Network Virtual Terminal (NVT) is an imaginary device consisting of a printer (or display) and a keyboard. Note that here the two peripherals play a contradictory role. The keyboard acts as an output device for producing outbound data to be sent over the TELNET connection.
- **Telnet Talk:** The DARPA telnet is an interactive communication protocol. It allows the remote session on the machine daemon service. It provides this service by typing the /usr/sbin/in.telnetd. It prompts you to enter the user name and the password.
- **AOL Mail:** AOL Mail, formalised as AOL Mail, is a free web-based email service provided by AOL, a division of Verizon Communications. In 1993, both America Online (AOL) and Delphi started connecting their proprietary e-mail services to the Internet.

5.20 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Question

1. Explain the basic concepts of Internet.
2. Define the importance of Internetworking?
3. What is meant by Internet and its groups of networks?
4. Elaborate on the term Domain Name System (DNS).
5. How are the TELNET commands performed?

6. State the use of e-mail in Internet Services.
7. Define MODEM and its use in Internet connection.
8. Briefly define all the components of Web Browser.
9. Differentiate between client side scripting and server side scripting.
10. List any two types of search engines along with their salient features.
11. Explain briefly about all the characteristics of NVT device.
12. What is Usenet?
13. What do you mean by AOL?

NOTES

Long-Answer Question

1. Explain briefly about the Internet Services along with the use of FTP, Archie, Finger and WWW.
2. Briefly describe the definition of e-mail along with its appropriate uses.
3. Elucidate the procedure of writing an e-mail message.
4. Describe about the use and operations of web browsers giving appropriate examples.
5. Discuss briefly about the equipment required for connecting to the Internet Services.
6. Explain the functioning of Webpage and its classification defining their salient features.
7. Elaborate the significance of home page explaining the steps of creation of a home page.
8. What do you mean by Index-Based Search Engine? Distinguish between Google and Yahoo Search Engines.
9. Describe briefly about TELNET protocol and its operations along with its modes of operations.
10. Elaborate the concept of AOL and discuss briefly about its features.

5.21 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

NOTES

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 6 ONLINE SOURCES OF DATA

NOTES

Structure

6.0 INTRODUCTION

A data source refers to the initial location where data originated or where physical information was first digitized, however even the most developed data may serve as a source, as long as another process accesses and utilizes it. Fundamentally, a data source may be a database, a flat file, and updated measurements from the specialized physical devices, scraped web data, or any of the numerous static which are available across the Internet. Databases remain the most common data sources, as the primary stores for data is in universal Relational DataBase Management Systems (RDBMS). In this perspective, the significant concept is the Data Source Name (DSN) which is defined within destination databases or applications as a pointer to the actual data, whether it exists locally or is found on a remote server, and whether in a single physical location or virtualized. The DSN is not necessarily the same as the relevant database name or file name, rather it is in an address or label used to easily reach the data at its source.

Though the diversity of content, format, and location for data is fast growing with contributions from technologies, such as IoT (Internet of Things) and the adoption of big data methodologies, hence it requires to classify the data sources. The data sources are classified into two broad categories as machine data sources and file data sources.

Nowadays, the Internet provides wide range of online data to its readers in the form of books, journals, working papers, reports, and newspapers. The books provide complete information (causes, effects, long-term consequences, fuller conclusions, etc.) for deeper analysis. Journals, newspapers, working papers and

reports include recent developments and events. Journals contain original researched information while the newspapers, working papers and reports may refer to researched studies done by others and hence it do not contain original information.

NOTES

In this unit, you will study about the basics of online sources of data, books, journals, working papers, reports, and newspapers.

6.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the significance of online sources of data
- Explain the types of source data
- Elaborate on Books, Journals, Working Papers, Reports and Newspapers

6.2 ONLINE SOURCES OF DATA

The Internet is referred as a very important source of online data that is available to all the Internet users, anytime and anywhere. Fundamentally, a data source is referred as the initial or original location where the data was originally originated or born, i.e., where the physical information was first digitized. However, the most developed data may serve as a source, provided that another process accesses and utilizes it. Basically, a data source may be a database, a flat file, and updated measurements from specified specialized physical devices, scraped web data, or any of the innumerable static which is available in abundance across the Internet.

The data on the Internet is typically stored and available in the form of databases. Because the databases are the most common data sources, they refer as the primary stores for data that is available and accessible everywhere follows the universal Relational DataBase Management Systems (RDBMS). In this perspective, the significant concept is included which is termed as the Data Source Name (DSN). The DSN is typically defined within destination databases or applications as a pointer to the actual data, whether it exists either locally or is found on a remote server, whether in a single physical location or virtualized. The DSN is simply an address or label specifically used to reach the data easily at its source and it may not essentially the equivalent as the relevant database name or file name.

Data Source Name (DSN)

In computing, a Data Source Name (DSN), sometimes known as a Database Source Name, though 'Data Sources' can comprise other repositories apart from databases, is a string that has an associated data structure used to describe a connection to a data source. Most commonly used in connection with ODBC

(Open DataBase Connectivity), DSNs also exist for JDBC (Java DataBase Connectivity) and for other data access mechanisms. The term often overlaps with 'Connection String'. Most systems do not make a distinction between DSNs or connection strings and the term can often be used interchangeably.

NOTES

DSN attributes may include, but are not limited to:

- The name of the data source.
- The location of the data source.
- The name of a database driver which can access the data source.
- A user ID for data access (if required).
- A user password for data access (if required).

The system administrator of a client machine generally creates a separate DSN for each relevant data source.

Standardizing DSNs offers a level of indirection, various applications, for example Apache/PHP and IIS/ASP, can take advantage of this in accessing shared data sources.

Types of Data Source

However, the diversity of content, format, and location for online data is fast growing/increasing with contributions from technologies, such as IoT (Internet of Things) and the adoption of big data methodologies, hence it requires to classify the data sources. The data sources are classified into two broad categories as machine data sources and file data sources.

Even though both the machine data sources and the file data sources share the same basic purpose, i.e., pointing to the location of data and defining analogous connection characteristics, hence machine data sources and file data sources can be stored, accessed, and used in different ways.

Types of Data Source Name (DNS)

The following two kinds of DSN exist:

- **Machine DSNs:** The Machine DNS is stored in collective configuration files (for example, /etc./odbc.ini, ~/.odbc.ini) and/or system resources (for example, Windows Registry HKLM\Software\ODBC\odbc.ini).
- **File DSNs:** The File DNS is stored in the file system with one DSN per file. File data sources contain all of the connection information inside a single, shareable, computer file (typically with a .dsn extension).

These are further categorised into the following two types:

- **System DSNs:** The System DNS is accessible by any and all processes and users of the system, stored in a centralized location (for example, /etc./ odbc.ini, /etc./odbc_file_dsns/<filename>).

- **User DSNs:** The User DNS is accessible only by the user who created the DSN, stored in a user-specific location (for example, ~/.odbc.ini, ~/odbc_file_dsns/<filename>).

NOTES

How Data Sources Work

Data sources can be used in various ways. Data can be transported using the diverse network protocols, such as the well-recognised File Transfer Protocol (FTP) and HyperText Transfer Protocol (HTTP), or any of the numerous Application Programming Interfaces (APIs) provided by the different authentic websites, networked applications, and other services.

Other protocols for moving data from sources to destinations, especially on the web, include NFS, SMB, SOAP, REST, and WebDAV. These protocols are typically used within APIs, within completely featured data applications, or as standalone transfer processes.

6.2.1 Books

A book is a medium for recording information in the form of writing or images, typically composed of many pages (made of papyrus, parchment, vellum, or paper) bound together and protected by a cover. The technical term for this physical arrangement is codex (plural, codices). Basically, a book is a self-sufficient section or part of a longer composition, a usage that reflects the fact that, in antiquity, long works had to be written on several scrolls, and each scroll had to be identified by the book it contained. So, for example, each part of Aristotle's Physics is called a book. Principally, a book is the compositional whole of which such sections, whether called books or chapters or parts, are parts.

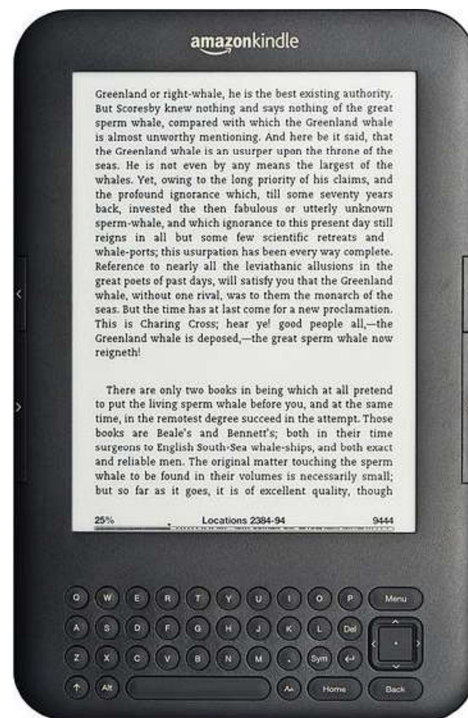
Books can consist only of drawings, engravings, or photographs, or such things as crossword puzzles or cut-out dolls. In a physical book, the pages can be left blank or can feature an abstract set of lines as support for ongoing entries, e.g., an account book, an appointment book, an autograph book, a notebook, a diary, or a sketchbook. Some physical books are made with pages thick and sturdy enough to support other physical objects, like a scrapbook or photograph album. Books may be distributed in electronic form as e-books and other formats.

E-BOOK

The term 'E-Book' is the abbreviation of 'Electronic Book'; it refers to a book-length publication in digital form. An E-Book is usually made available through the Internet, but also on CD-ROM and other forms. In the 2000s, due to the rise in availability of affordable handheld computing devices, the opportunity to share texts through electronic means became an interesting option for media publishers, consequently the 'E-Book' originated. E-Books may be read either via a computing device with an LED display, such as a traditional computer, a smartphone or a tablet computer; or by means of a portable e-ink display device known as an 'E-

NOTES

Book Reader', such as the Sony Reader, Barnes & Noble Nook, Kobo eReader, or the Amazon Kindle. E-Book readers attempt to mimic the experience of reading a print book by using this technology, since the displays on E-Book readers are much less reflective. The following screen illustrates a Kindle E-Reader, a form of E-Book.



Definition 1: An Electronic Book, also known as an E-Book or eBook, is a book publication made available in digital form, consisting of text, images, or both, readable on the flat-panel display of computers or other electronic devices. Although sometimes defined as 'An Electronic Version of a Printed Book', some E-Books exist without a printed equivalent. E-Books can be read on dedicated 'E-Reader Devices', but also on any computer device that features a controllable viewing screen, including desktop computers, laptops, tablets and smartphones.

Definition 2: E-Books are also referred to as 'ebooks', 'eBooks', 'Ebooks', 'e-Books', 'e-journals', 'e-editions', or 'digital books'. A device that is designed specifically for reading e-books is called an 'e-reader', 'ebook device', or 'eReader'.

Consequently, in the 2000s, there was a trend of publishing the E-Book on the Internet, where readers used to buy traditional paper books and also E-Books on websites using e-commerce systems. With e-books, users can browse through titles online, and then when they select and order titles, the E-Book can be sent to them online or the user can download the E-Book. The key reasons for people buying E-Books are probably the lower prices and improved comfort as they can buy the required E-Books from home or on the go with mobile devices.

E-Book Formats

NOTES

The specific format of E-Book was supported from major software companies, such as Adobe with its PDF format that was introduced in 1993. Unlike most other formats, PDF documents are generally tied to a particular dimension and layout, rather than adjusting dynamically to the current page, window, or another size. Different e-reader devices followed different formats, most of them accepting books in only one or a few formats.

The specific approach termed as Text Encoding Initiative (TEI) was formed and developed as typical guidelines for encoding books and other materials of interest for a variety of analytic uses as well as reading, and countless literary and other works have been developed using this TEI approach. In the late 1990s, a consortium formed to develop the Open eBook format as a way for authors and publishers to provide a single source-document which many book-reading software and hardware platforms could handle. Several specialists from the TEI were closely involved in the early development of Open eBook. Focused on portability, Open eBook as defined required subsets of XHTML and CSS; a set of multimedia formats, and an XML schema for a 'manifest', to list the components of a given e-book, identify a table of contents, cover art, and so on. Google Books has converted many public domain works to this open format. Many E-Book publishers started distributing books that were in the public domain.

Depending on possible digital rights management, the E-Books can be backed up and recovered in the case of loss or damage to the device on which they are stored, a new copy can be downloaded without incurring an additional cost from the distributor, as well as being able to synchronize the reading location, highlights and bookmarks across several devices.

Public Domain Books

Public domain books are those whose copyrights have expired, meaning they can be copied, edited, and sold freely without restrictions. Many of these books can be downloaded for free from websites like the Internet Archive, in formats that many e-readers support, such as PDF, TXT, and EPUB. Books in other formats may be converted to an e-reader-compatible format using E-Book writing software, for example Calibre.

6.2.2 Journals

A 'Journal', from the Old French journal meaning 'Daily', may refer to several things. In its original meaning, it refers to a daily record of activities, but the term has evolved to mean any record of activities, regardless of time elapsed between entries, such as a,

- Diary, a record of what happened over the course of a day or other period.
- Daybook, also known as general journal, a daily record of financial transactions.

- Logbook, a record of events important to the operation of a vehicle, facility, or otherwise.
- Transaction journal, a chronological record of data processing.

In publishing, it can also refer to various periodicals or serials:

- Academic journal, an academic or scholarly periodical.
  - o Scientific journal, an academic journal focusing on science
  - o Medical journal, an academic journal focusing on medicine
- Magazine, non-academic or scholarly periodicals in general.
  - o Trade magazine, a magazine of interest to those of a particular profession or trade.
  - o Literary magazine, a magazine devoted to literature in a broad sense.
- Newspaper, a periodical covering general news and current events in politics, business, sports and art.
  - o Gazette, a type of newspaper, often a newspaper of record.
  - o Government gazette, a government newspaper which publishes public or legal notices.

Electronic Journal

Electronic Journals, also known as ejournals, e-journals, and electronic serials, are scholarly journals or intellectual magazines that can be accessed via electronic transmission.

Some journals are 'Born Digital' in that they are solely published on the web and in a digital format, but most electronic journals originated as print journals, which subsequently evolved to have an electronic version, while still maintaining a print component. With the growth of the Internet, the e-journal has come to dominate the journals world.

An e-journal closely resembles a print journal in structure, as there is a table of contents which lists the articles, and many electronic journals still use a volume/ issue model, although some titles now publish on a continuous basis. Online journal articles are a specialized form of electronic document as they have the purpose of providing material for academic research and study, and they are formatted approximately like journal articles in traditional printed journals. Often a journal article will be available for download in two formats - as a PDF and in HTML format, although other electronic file types are often supported for supplementary material. Articles are indexed in bibliographic Databases, as well as by Search Engines. E-journals permit new types on content to be included in journals, for example video material, or the data sets on which research has been based.

With the growth and development of the Internet, there has been a growth in the number of new journals, especially in those that exist as digital publications

NOTES

only. A subset of these journals exist as Open Access titles, meaning that they are free to access for all, and have Creative Commons licences which permit the reproduction of content in different ways. High quality open access journals are listed in Directory of Open Access Journals. Most however continue to exist as subscription journals.

6.2.3 Working Papers

Working Papers are defined as the pre-publication versions of academic articles, book chapters, or reviews. Working Papers are not refereed in any document as they need authentication. The form and content of Working Papers are the responsibility of individual authors.

A working paper or work paper may be defined as:

1. A preliminary scientific or technical paper. Often, authors will release working papers to share ideas about a topic or to elicit feedback before submitting to a peer reviewed conference or academic journal.
2. Working papers are often the basis for related works, and may in themselves be cited by peer-review papers. They may be considered as grey literature.

Sometimes the term working paper is used synonymously as technical report. Working papers are typically hosted on websites, belonging either to the author or the author's affiliated institution, for example the United Nations uses the term 'Working Paper' in approximately this sense for the draft of a resolution.

6.2.4 Reports

A report is a document that presents information in an organized format for a specific viewers and purpose. Although summaries of reports may be delivered orally, complete reports are almost always in the form of written documents.

Definition: A 'Report' can be defined as the document containing information organized in a narrative, graphic, or tabular form, prepared on ad hoc, periodic, recurring, regular, or as required basis. Reports may refer to specific periods, events, occurrences, or subjects, and may be communicated or presented in oral or written form to provide information.

In modern scenario, reports play a key role in the progress of any organization. Reports are the backbone to the thinking process of the establishment and they are responsible, to a great extent, in evolving an efficient or inefficient work environment.

Following are the significance of the reports:

- Reports present adequate information on various aspects.
- All the significant information are communicated through reports.
- Reports help in decision-making.

NOTES

- A rule and balanced report also helps in problem solving.
- Reports communicate the planning, policies and other matters regarding an organization to the masses.
- News reports play the role of ombudsman and levy checks and balances on the establishment.

One of the most common formats for presenting reports is 'IMRAD—Introduction, Methods, Results, And Discussion'. This structure is the standard for the genre, mirrors traditional publication of scientific research and summons the ethos and credibility of that discipline. Generally, 'Reports' are not required to follow this pattern and may use alternative methods, such as the problem-solution format, wherein an issue is first listed followed by details that what must be done to fix the problem. Transparency and a focus on quality are keys to writing a useful report. Accuracy is also significant.

Standard Elements of Reports

Reports use typical features, such as tables, graphics, pictures, voice, or specialized vocabulary in order to persuade a specific viewers to undertake an action or inform the reader of the subject. Some common elements of written reports include headings to indicate topics and help the reader locate relevant information quickly, and visual elements, such as charts, tables and figures, which are useful for breaking up large sections of text and making complex issues more accessible. Lengthy written reports will almost always contain a table of contents, appendices, footnotes, and references. A bibliography or list of references will appear at the end of any credible report and citations are often included within the text itself. Complex terms are explained within the body of the report or listed as footnotes in order to make the report easier to follow. A short summary of the report's contents, called an abstract, may appear in the beginning so that the readers/viewers knows what the report will cover. Online reports often contain hyperlinks to internal or external sources as well.

Quality reports will be well researched and the presenter will list their sources if at all possible.

Types of Reports

Following are some examples of reports:

- Annual Reports
- Auditor's Reports
- Book Reports
- Bound Report
- Retail Report
- Census Reports

NOTES

- Credit Reports
- Demographic Reports
- Expense Report
- Experience Report
- Inspection Reports
- Military Reports
- Police Reports
- Policy Reports
- Informal Reports
- Progress Reports
- Investigative Reports
- Technical or Scientific Reports
- Trip Reports
- White Papers
- Appraisal Reports
- Workplace Reports

A report is typically written for a clear purpose and to a particular spectators/ readers. Specific information and evidence are presented, analysed and applied to a particular problem or issue. The information is presented in a clearly structured format making use of sections and headings so that the information is easy to discover. A well-written report usually provides the clear instructions and guidelines.

What is a Good Report?

Following are the two reasons why reports are used as forms of written assessment:

- To find out what you have learned from your reading, research or experience.
- To give you experience of an important skill that is widely used in the work place.

An effective report presents and analyses facts and evidence that are relevant to the specific problem or issue of the report brief. All sources used should be acknowledged and referenced throughout, in accordance with the preferred method.

Plagiarism must be avoided.

A well written report will establish the ability to:

- Understand the purpose of the report brief and adhere to its specifications.

NOTES

- Gather, evaluate and analyse relevant information.
- Structure material in a logical and coherent order.
- Present your report in a consistent manner according to the instructions of the report brief.
- Make appropriate conclusions that are supported by the evidence and analysis of the report.
- Make thoughtful and practical recommendations where required.

The structure of a report includes the Title Page, Contents (Table of Contents), Introduction, Methods, Results, Conclusion or Summary (Abstract), Terms of Reference, Appendices, Bibliography, Acknowledgements, and Glossary of Technical Terms.

6.2.5 Newspapers

A newspaper is a periodical publication containing written information about current events and is often typed in black ink with a white or grey background.

Newspapers developed in the 17th century, as information sheets for merchants. By the early 19th century, many cities in Europe, as well as North and South America, published newspapers.

Some newspapers with high editorial independence, high journalism quality, and large circulation are viewed as newspapers of record.

Definition: Newspaper, publication usually issued daily, weekly, or at other regular times that provides news, views, features, and other information of public interest and that often carries advertising.

Newspapers can cover a wide variety of fields, such as politics, business, sports and art, and often include materials, such as opinion columns, weather forecasts, and reviews of local services, obituaries, birth notices, crosswords, editorial cartoons, comic strips, and advice columns. Most newspapers are businesses, and they pay their expenses with a mixture of subscription revenue, newsstand sales, and advertising revenue. The journalism organizations that publish newspapers are themselves often metonymically called newspapers.

Newspapers have traditionally been published in print (usually on cheap, low-grade paper called newsprint). However, today most newspapers are also published on websites as ‘Online Newspapers’, and some have even abandoned their print versions entirely.

Online Newspaper

An ‘Online Newspaper’ is the online version of a newspaper, either as a stand-alone publication or as the online version of a printed periodical.

NOTES

Online Newspapers have created more opportunities for newspapers, such as competing with broadcast journalism in presenting breaking news in appropriate timely manner. The credibility and strong brand recognition of well-established newspapers, and the close relationships they have with advertisers, are also seen by many in the newspaper industry as strengthening their chances of survival. The movement away from the printing process can also help decrease costs.

Online Newspapers, like printed newspapers, have legal restrictions regarding libel, privacy, and copyright, also apply to online publications in most countries.

News reporters are being specifically taught to shoot video and to write in the concise manner essentially for Internet news pages. Some newspapers have attempted to integrate the Internet into every aspect of their operations, for example, the writing of stories for both print and online, and classified advertisements appearing in both media, while other newspaper websites may be quite different from the corresponding printed newspaper.

Newspapers are typically published daily or weekly. News magazines are also weekly, but they have a magazine format. General-interest newspapers typically publish news articles and feature articles on national and international news as well as local news. The news includes political events and personalities, business and finance, crime, weather, and natural disasters; health and medicine, science, and computers and technology; sports; and entertainment, society, food and cooking, clothing and home fashion, and the arts.

Most traditional papers also feature an editorial page containing editorials written by an editor (or by the paper's editorial board) and expressing an opinion on a public issue, opinion articles called 'Op-Eds' written by guest writers (which are typically in the same section as the editorial), and columns that express the personal opinions of columnists, usually offering analysis and synthesis that attempts to translate the raw data of the news into information telling the reader 'What It All Means'.

A wide variety of material has been published in the Newspapers. Besides the aforementioned news, information and opinions, they include weather forecasts; criticism and reviews of the arts (including literature, film, television, theatre, fine arts, and architecture) and of local services, such as restaurants; obituaries, birth notices and graduation announcements; entertainment features, such as crosswords, horoscopes, editorial cartoons, gag cartoons, and comic strips; advice columns, food, and other columns; and radio and television listings (program schedules). As of 2017, newspapers may also provide information about new movies and TV shows available on streaming video services like 'Netflix'. Newspapers have classified ad sections where people and businesses can buy small advertisements to sell goods or services; as of 2013, the huge increase in Internet websites for selling goods, such as Craigslist and eBay has led to significantly less classified ad sales for newspapers.

NOTES

Newspapers typically observe to the following four criteria:

Public Accessibility: The contents of the Newspapers are reasonably accessible to the public, traditionally by the paper being sold or distributed at newsstands, shops, and libraries, and, since the 1990s, made available over the Internet with online newspaper websites. While online newspapers have increased access to newspapers by people with Internet access, people without Internet or computer access will not be able to read online news.

Periodicity: The Newspapers are published at regular intervals, typically daily or weekly. This ensures that newspapers can provide information on newly-emerging news stories or events.

Currency: The information given in the Newspapers is up-to-date as its publication schedule permits. The degree of up-to-date-ness of a print newspaper is limited by the need of time to print and distribute the newspaper. Online newspapers can be updated as frequently as new information becomes available, even a number of times per day, which means that online editions can be more appropriate and up-to-date.

Universality: Newspapers covers a range of topics, from political and business news to updates on science and technology, arts, culture, and entertainment.

Check Your Progress

1. Why the Internet is referred as a very important source of online data?
2. What is a data source?
3. Define the term Data Source Name (DSN).
4. How the data from the data sources can be transported on the Internet?
5. Explain the term E-Book. How the E-Books can be read?

6.3 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The Internet is referred as a very important source of online data because it is available to all the Internet users, anytime and anywhere.
2. A data source is referred as the initial or original location where the data was originally originated or born, i.e., where the physical information was first digitized. Basically, a data source may be a database, a flat file, and updated measurements from specified specialized physical devices, scraped web data, or any of the innumerable static which is available in abundance

NOTES

across the Internet. The data on the Internet is typically stored and available in the form of databases. Because the databases are the most common data sources, they refer as the primary stores for data that is available and accessible everywhere follows the universal Relational DataBase Management Systems (RDBMS).

3. The Data Source Name (DSN) is typically defined within destination databases or applications as a pointer to the actual data, whether it exists either locally or is found on a remote server, whether in a single physical location or virtualized. The DSN is simply an address or label specifically used to reach the data easily at its source and it may not essentially the equivalent as the relevant database name or file name.
4. Data sources can be used in various ways. Data can be transported using the diverse network protocols, such as the well-recognised File Transfer Protocol (FTP) and HyperText Transfer Protocol (HTTP), or any of the numerous Application Programming Interfaces (APIs) provided by the different authentic websites, networked applications, and other services. Other protocols for moving data from sources to destinations, especially on the web, include NFS, SMB, SOAP, REST, and WebDAV. These protocols are typically used within APIs, within completely featured data applications, or as standalone transfer processes.
5. The term 'E-Book' is the abbreviation of 'Electronic Book'; it refers to a book-length publication in digital form. An E-Book is usually made available through the Internet, but also on CD-ROM and other forms. An Electronic Book, also known as an E-Book or eBook, is a book publication made available in digital form, consisting of text, images, or both, readable on the flat-panel display of computers or other electronic devices. E-Books can be read on dedicated 'E-Reader Devices', but also on any computer device that features a controllable viewing screen, including desktop computers, laptops, tablets and smartphones.

   E-Books are also referred to as 'ebooks', 'eBooks', 'Ebooks', 'e-Books', 'e-journals', 'e-editions', or 'digital books'. A device that is designed specifically for reading e-books is called an 'e-reader', 'ebook device', or 'eReader'.

6.4 SUMMARY

- The Internet is referred as a very important source of online data that is available to all the Internet users, anytime and anywhere.
- Fundamentally, a data source is referred as the initial or original location where the data was originally originated or born, i.e., where the physical

information was first digitized. However, the most developed data may serve as a source, provided that another process accesses and utilizes it.

NOTES

- Basically, a data source may be a database, a flat file, and updated measurements from specified specialized physical devices, scraped web data, or any of the innumerable static which is available in abundance across the Internet.
- The Data Source Name (DSN) is typically defined within destination databases or applications as a pointer to the actual data, whether it exists either locally or is found on a remote server, whether in a single physical location or virtualized.
- The DSN is simply an address or label specifically used to reach the data easily at its source and it may not essentially the equivalent as the relevant database name or file name.
- In computing, a Data Source Name (DSN), sometimes known as a Database Source Name, though 'Data Sources' can comprise other repositories apart from databases, is a string that has an associated data structure used to describe a connection to a data source.
- Most commonly used in connection with ODBC (Open DataBase Connectivity), DSNs also exist for JDBC (Java DataBase Connectivity) and for other data access mechanisms. The term often overlaps with 'Connection String'.
- The diversity of content, format, and location for online data is fast growing/ increasing with contributions from technologies, such as IoT (Internet of Things) and the adoption of big data methodologies, hence it requires to classify the data sources. The data sources are classified into two broad categories as machine data sources and file data sources.
- The Machine DNS is stored in collective configuration files (for example, / etc./odbc.ini, ~/.odbc.ini) and/or system resources (for example, Windows Registry HKLM\Software\ODBC\odbc.ini).
- The File DNS is stored in the file system with one DSN per file. File data sources contain all of the connection information inside a single, shareable, computer file (typically with a .dsn extension).
- The System DNS is accessible by any and all processes and users of the system, stored in a centralized location (for example, /etc./odbc.ini, /etc./ odbc_file_dsns/<filename>).
- The User DNS is accessible only by the user who created the DSN, stored in a user-specific location (for example, ~/.odbc.ini, ~/odbc_file_dsns/ <filename>).
- Data sources can be used in various ways. Data can be transported using the diverse network protocols, such as the well-recognised File Transfer

Protocol (FTP) and HyperText Transfer Protocol (HTTP), or any of the numerous Application Programming Interfaces (APIs) provided by the different authentic websites, networked applications, and other services.

NOTES

- Other protocols for moving data from sources to destinations, especially on the web, include NFS, SMB, SOAP, REST, and WebDAV. These protocols are typically used within APIs, within completely featured data applications, or as standalone transfer processes.
- A book is a medium for recording information in the form of writing or images, typically composed of many pages (made of papyrus, parchment, vellum, or paper) bound together and protected by a cover. Principally, a book is the compositional whole of which such sections, whether called books or chapters or parts, are parts.
- The term 'E-Book' is the abbreviation of 'Electronic Book'; it refers to a book-length publication in digital form. An E-Book is usually made available through the Internet, but also on CD-ROM and other forms.
- An Electronic Book, also known as an E-Book or eBook, is a book publication made available in digital form, consisting of text, images, or both, readable on the flat-panel display of computers or other electronic devices. Although sometimes defined as 'An Electronic Version of a Printed Book', some E-Books exist without a printed equivalent. E-Books can be read on dedicated 'E-Reader Devices', but also on any computer device that features a controllable viewing screen, including desktop computers, laptops, tablets and smartphones.
- E-Books are also referred to as 'ebooks', 'eBooks', 'Ebooks', 'e-Books', 'e-journals', 'e-editions', or 'digital books'. A device that is designed specifically for reading e-books is called an 'e-reader', 'ebook device', or 'eReader'.
- A 'Journal', from the Old French journal meaning 'Daily', may refer to several things. In its original meaning, it refers to a daily record of activities, but the term has evolved to mean any record of activities, regardless of time elapsed between entries.
- Electronic Journals, also known as ejournals, e-journals, and electronic serials, are scholarly journals or intellectual magazines that can be accessed via electronic transmission.
- Working Papers are defined as the pre-publication versions of academic articles, book chapters, or reviews. Working Papers are not refereed in any document as they need authentication. The form and content of Working Papers are the responsibility of individual authors.

NOTES

- A report is a document that presents information in an organized format for a specific viewers and purpose. Although summaries of reports may be delivered orally, complete reports are almost always in the form of written documents.
- A 'Report' can be defined as the document containing information organized in a narrative, graphic, or tabular form, prepared on ad hoc, periodic, recurring, regular, or as required basis. Reports may refer to specific periods, events, occurrences, or subjects, and may be communicated or presented in oral or written form to provide information.
- A report is typically written for a clear purpose and to a particular spectators/readers. Specific information and evidence are presented, analysed and applied to a particular problem or issue. The information is presented in a clearly structured format making use of sections and headings so that the information is easy to discover. A well-written report usually provides the clear instructions and guidelines.
- The structure of a report includes the Title Page, Contents (Table of Contents), Introduction, Methods, Results, Conclusion or Summary (Abstract), Terms of Reference, Appendices, Bibliography, Acknowledgements, and Glossary of Technical Terms.
- A newspaper is a periodical publication containing written information about current events and is often typed in black ink with a white or grey background.
- Newspaper, publication usually issued daily, weekly, or at other regular times that provides news, views, features, and other information of public interest and that often carries advertising.
- Newspapers can cover a wide variety of fields, such as politics, business, sports and art, and often include materials, such as opinion columns, weather forecasts, and reviews of local services, obituaries, birth notices, crosswords, editorial cartoons, comic strips, and advice columns.
- An 'Online Newspaper' is the online version of a newspaper, either as a stand-alone publication or as the online version of a printed periodical.
- Online newspapers can be updated as frequently as new information becomes available, even a number of times per day, which means that online editions can be more appropriate and up-to-date.

6.5 KEY WORDS

- **Data source:** The data source is referred as the initial or original location where the data was originally originated or born, i.e., where the physical information was first digitized. Basically, a data source may be a database,

NOTES

a flat file, and updated measurements from specified specialized physical devices, scraped web data, or any of the innumerable static which is available in abundance across the Internet.

- **Data Source Name (DSN):** The Data Source Name (DSN) is typically defined within destination databases or applications as a pointer to the actual data, whether it exists either locally or is found on a remote server, whether in a single physical location or virtualized. The DSN is simply an address or label specifically used to reach the data easily at its source and it may not essentially the equivalent as the relevant database name or file name.
- **Machine DSNs:** The Machine DNS is stored in collective configuration files (for example, /etc./odbc.ini, ~/.odbc.ini) and/or system resources (for example, Windows Registry HKLM\Software\ODBC\odbc.ini).
- **File DSNs:** The File DNS is stored in the file system with one DSN per file. File data sources contain all of the connection information inside a single, shareable, computer file (typically with a .dsn extension).
- **System DSNs:** The System DNS is accessible by any and all processes and users of the system, stored in a centralized location (for example, /etc./odbc.ini, /etc./odbc_file_dsns/<filename>).
- **User DSNs:** The User DNS is accessible only by the user who created the DSN, stored in a user-specific location (for example, ~/.odbc.ini, ~/odbc_file_dsns/<filename>).
- **E-Book:** An Electronic Book, also known as an E-Book or eBook, is a book publication made available in digital form, consisting of text, images, or both, readable on the flat-panel display of computers or other electronic devices.
- **Electronic Journals:** The Electronic Journals, also known as ejournals, e-journals, and electronic serials, are scholarly journals or intellectual magazines that can be accessed via electronic transmission.
- **Report:** A report is a document that presents information in an organized format for a specific viewers and purpose, i.e., the document containing information organized in a narrative, graphic, or tabular form, prepared on ad hoc, periodic, recurring, regular, or as required basis.

6.6 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is a data source?
2. How is data stored in the data source?

NOTES

3. Differentiate between Books and E-Books.
4. Explain the term Journals and E-Journals.
5. What are Working Papers?
6. Why the Reports are made?
7. What is the importance of Newspapers?

Long-Answer Questions

1. Briefly discuss the significance of 'Online Sources of Data' giving appropriate examples.
2. Why the Internet is referred as a very important source of online data? Support your answer with the help of appropriate examples.
3. Discuss the concept and significance of Relational DataBase Management Systems (RDBMS) and the Data Source Name (DSN) for working with online data sources. Support your answer with the help of appropriate examples.
4. Explain the various types of data source giving suitable examples.
5. Discuss the basic concept of Books and E-Books giving the details of relevant features, format and public domain books.
6. Write short notes on the following:
   (i) Journals
   (ii) Working Papers
   (iii) Reports
   (iv) Newspapers

6.7 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

NOTES

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 7 BASICS OF MS-OFFICE AND MS-WORD 2007

Structure

NOTES

7.0 INTRODUCTION

Microsoft Office 2007 (codenamed Office 12) is a version of Microsoft Office, a family of office suites and productivity software for Windows, developed and published by Microsoft. It was preceded by Office 2003 and succeeded by Office 2010. Office 2007 introduced a new Graphical User Interface (GUI) called the Fluent User Interface (FUI), which uses ribbons and an Office menu instead of menu bars and toolbars. Office 2007 also introduced Office Open XML file formats as the default file formats in MS Excel, MS PowerPoint, and MS Word. The new formats are intended to facilitate the sharing of information between programs, improve security, reduce the size of documents, and enable new recovery scenarios.

Microsoft Word 2007 is a full featured word processing program, which can be used for any work involving creating and managing text—write letters and reports, prepare bills, invoices, office stationery (letter-heads, envelopes, forms, etc.), design brochures, pamphlets, newsletters, magazines, etc. The 2007 version of the program has introduced the 'Ribbon User Interface' which is a work oriented Graphical User Interface or GUI. The Ribbon is positioned at the top of the screen of the Word Window and includes eight tabs, such as Home, Insert, Page Layout, References, Mailings, Review, View and Add-Ins. You will learn to create a document, save a document, send a document by e-mail or fax, format a document as per your requirement and print it. In Word 2007, the AutoCorrect, thesaurus and spell checking options help you to check spellings and grammar in a document. Word 2007 also provides feature to insert/create table in a document for displaying data in a tabular format and to sort data in ascending or descending order. The mail merge feature is the prime application of word processing technology because this tool allows you to send personalized messages to a large group of people at once without them knowing who else you have sent it to.

In this unit, you will study about the basics of MS-Office, MS-Word, the MS-Word Window, entering, selecting, copying and moving text, applying fonts and indenting text, creating numbering and bullets, finding and replacing text, spelling, grammar and thesaurus, creating page headers and footers, constructing tables and Mail Merge features.

7.1 OBJECTIVES

NOTES

After going through this unit, you will be able to:

- Discuss the significance of MS Office 2007.
- Describe the significant features of Microsoft Word 2007
- Understand the significance of the Microsoft Office button and Ribbon
- Work with multiple documents and describe various types of document views
- Customize the Word environment and edit a document using various Word features
- Add tables, symbols, special characters and equations
- Proofread the document using Spelling & Grammar, thesaurus and AutoCorrect
- Insert and resize illustrations, pictures, Clip Art and SmartArt graphics
- Use mail merge feature to send messages to a large group of people

7.2 BASICS OF MS-OFFICE 2007

Microsoft Office 2007 (codenamed Office 12) is a version of Microsoft Office, a family of office suites and productivity software for Windows, developed and published by Microsoft. It was preceded by Office 2003 and succeeded by Office 2010.



Office 2007 introduced a new Graphical User Interface (GUI) called the Fluent User Interface (FUI), which uses ribbons and an Office menu instead of menu bars and toolbars. Office 2007 also introduced Office Open XML file formats as the default file formats in MS Excel, MS PowerPoint, and MS Word. The new formats are intended to facilitate the sharing of information between programs, improve security, reduce the size of documents, and enable new recovery scenarios.

Office 2007 requires Windows XP with Service Pack 2, Windows Server 2003 with Service Pack 1, or a later version of Windows; it is the last version of Microsoft Office to run on Windows XP Professional x64 Edition.

Office 2007 includes new applications and server-side tools, such as Office SharePoint Server 2007, a major revision to the server platform for Office applications, which supports Excel Services, a client-server architecture for supporting Excel workbooks that are shared in real time between multiple machines, and are also viewable and editable through a web page. With Microsoft FrontPage

NOTES

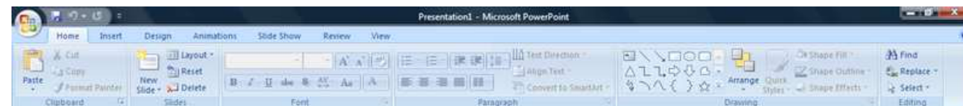
discontinued, Microsoft SharePoint Designer, which is aimed towards development of SharePoint portals, becomes part of the Office 2007 family.

Features of Microsoft Office 2007

User Interface: The new User Interface (UI), officially known as Fluent User Interface (FUI), has been implemented in the core Microsoft Office applications: Word, Excel, PowerPoint, Access, and Outlook. The default font used in this edition is Calibri.

Office Button: The Office 2007 button, located on the top-left of the window, replaces the File menu and provides access to functionality common across all Office applications, including opening, saving, printing, and sharing a file. It can also close the application. Users can also choose colour schemes for the interface.

Ribbon: The ribbon, a panel that houses a fixed arrangement of command buttons and icons, organizes commands as a set of tabs, each grouping relevant commands. The ribbon is present in Microsoft Word 2007, Microsoft Excel 2007, Microsoft PowerPoint 2007, Microsoft Access 2007 and some Microsoft Outlook 2007 windows. The ribbon is not user customizable in Office 2007. Each application has a different set of tabs that exposes functions that the application offers. For example, while Excel has a tab for the graphing capabilities, Word does not; instead it has tabs to control the formatting of a text document. Within each tab, various related options may be grouped together. The ribbon is designed to make the features of the application more discoverable and accessible with fewer mouse clicks as compared to the menu-based UI used prior to Office 2007. Moving the mouse scroll wheel while on any of the tabs on the ribbon cycles—through the tabs. The ribbon can be minimized by double clicking the active section's title, as shown in the 'Ribbon' in Microsoft PowerPoint 2007 Home text in the following screenshot.



Contextual Tabs: Some tabs, called Contextual Tabs, appear only when certain objects are selected. Contextual Tabs expose functionality specific only to the object with focus. For example, selecting a picture brings up the Pictures tab, which presents options for dealing with the picture. Similarly, focusing on a table exposes table-related options in a specific tab. Contextual Tabs remain hidden except when an applicable object is selected.

Live Preview: Microsoft Office 2007 also introduces a feature called Live Preview, which temporarily applies formatting on the focused text or object when on any formatting button the mouse is hovered. The temporary formatting is removed when the mouse pointer is moved from the button. This allows users to have a preview of how the option would affect the appearance of the object, without actually applying it.

Mini Toolbar: The new Mini Toolbar is a small toolbar with basic formatting commands that appears within the document editing area, much like a context menu. When the mouse selects part of the text, Mini Toolbar appears close to selected

NOTES

text. It remains semi-transparent until the mouse pointer is hovered on it, to avoid obstructing what is underneath. Mini Toolbar can also be made to appear by right-clicking in the editing area or via **'c" Menu'** key on keyboard, in which case it appears near the cursor, above or below the traditional context menu. Mini Toolbar is not customizable in Office 2007, but can be turned off.

Quick Access Toolbar: The Quick Access toolbar (by default) is positioned in the title bar and serves as a repository of most used functions, such as save, undo/redo and print. It is customizable, although this feature is limited, compared to toolbars in previous Office versions. Any command available in the entire Office application can be added to the Quick Access toolbar, including commands not available on the ribbon as well as macros. Keyboard shortcuts for any of the commands on the toolbar are also fully customizable, similar to previous Office versions.

Other User Interface (UI) Features

- Super-tooltips, or screentips, that can house formatted text and even images, are used to provide detailed descriptions of what most buttons do.
- A zoom slider present in the bottom-right corner, allowing for dynamic and rapid magnification of documents.
- The status bar is fully customizable. Users can right click the status bar and add or remove what they want the status bar to display.

SmartArt: SmartArt, found under the Insert tab in the ribbon in Microsoft Word 2007, Microsoft Excel 2007, Microsoft PowerPoint 2007, Microsoft Access 2007 and some Microsoft Outlook 2007 windows is a new group of editable and formatted diagrams. There are 115 preset SmartArt graphics layout templates in categories, such as list, process, cycle, and hierarchy. When an instance of a SmartArt is inserted, a Text Pane appears next to it to guide the user through entering text in the hierarchical levels. Each SmartArt graphic, based on its design, maps the text outline, automatically resized for best fit, onto the graphic. There are a number of 'Quick Styles' for each graphic that apply largely different 3D effects to the graphic, and the graphic's shapes and text can be formatted through shape styles and WordArt styles. In addition, SmartArt graphics change their colours, fonts, and effects to match the document's theme.

File Formats: Office Open XML

Microsoft Office 2007 introduced a new file format, called Office Open XML, as the default file format. Such files are saved using an extra X letter in their extension, such as **.docx** for Microsoft Word 2007, **.xlsx** for Microsoft Excel 2007, **.pptx** for Microsoft PowerPoint 2007, etc. However, it can still save documents in the old format, which is compatible with previous versions. Alternatively, Microsoft has made available a free add-on known as the Microsoft Office Compatibility Pack that lets Office 2000, XP, and 2003 open, edit, and save documents created under the newer 2007 format.

Office Open XML is based on XML and uses the ZIP file container. According to Microsoft, documents created in this format are up to 75% smaller

than the same documents saved with previous Microsoft Office file formats, owing to the ZIP data compression. Files containing macros are saved with an extra 'm' letter in their extension, such as .docm, xlsm, pptm, etc.

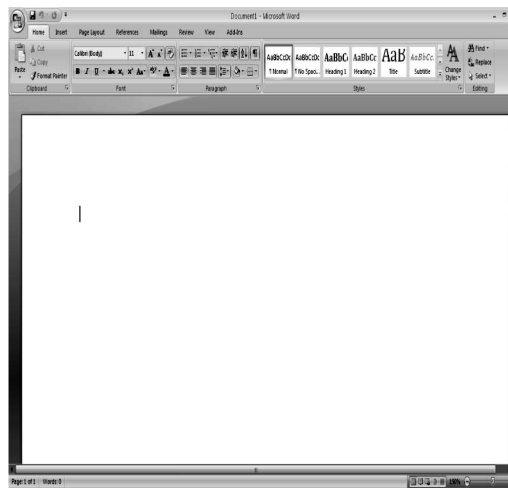
Portable Document Format (PDF): Initially, Microsoft promised to support exporting to Portable Document Format (PDF) in Office 2007. However, due to legal objections from Adobe Systems, Office 2007 originally did not offer PDF support out of the box, but rather as a separate free download.

7.3 INTRODUCTION TO MICROSOFT WORD 2007

Microsoft Word 2007 is a product of Microsoft Office 2007 or 2007 Microsoft Office System. It is the latest version of Windows of the Microsoft Office System. It introduced the 'Ribbon User Interface' instead of the old menu and toolbar which is a work oriented Graphical User Interface (GUI). You can create newsletters, articles, annual reports or update a blog using various formats and themes of Microsoft Word 2007. Themes help to change the presentation of a document.

7.3.1 Getting Started

The Start button in the lower left corner of your screen gives you access to all the programs on your PC and also to Word. To start Microsoft Word, Click on the **Start** button and select **All Programs.** To open this window, you will need to perform the following steps:



- Click on the **Start** button and select **Microsoft Office** from **All Programs.**
- Select **Microsoft Office Word 2007**.

  The user interface of Microsoft Word 2007 is shown in the screen.

Menus

When you explore Word 2007, you will notice the new look of the menu bar. Three new features help you to work with Word 2007, namely the Microsoft Office Button, the Quick Access Toolbar and the Ribbon which contain various functions.

The Microsoft Office Button

The Microsoft Office button is located in the upper left corner of the Word 2007 Window. A menu appears when you click on this button. This menu helps in creating a new document or file, opening an existing document or file, saving a document or file, printing a document or file, sending the document or file via fax or e-mail, etc.

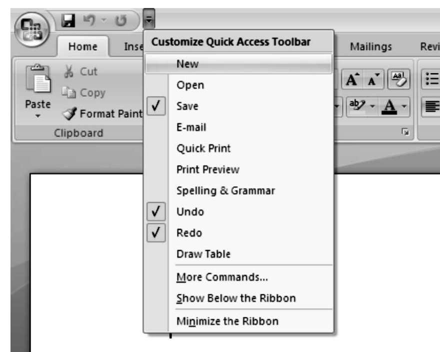


The Quick Access Toolbar

The Quick Access toolbar is right next to the Microsoft Office button. This toolbar helps you to access the frequently used commands. The default commands which appear on this toolbar are Save, Undo and Redo. These commands help you to Save a document or file, Undo an action and Redo an action.

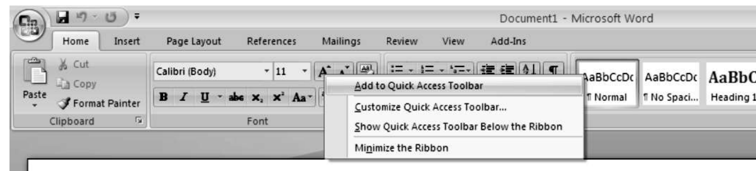


You can customize this toolbar as per your requirements by clicking on the expansion button as shown here.



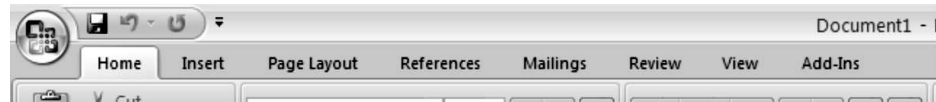
More items can be added to the quick access toolbar by right clicking on the item which you want to add in the Office Button or the Ribbon and then clicking on Add to Quick Access Toolbar as shown.

NOTES



The Ribbon

The Ribbon is positioned at the top of the screen of the Word window. It includes eight tabs, namely Home, Insert, Page Layout, References, Mailings, Review, View and Add-Ins as shown in screen below. Each tab contains various new and advanced features of Word.



Each tab specifically contains certain tools as follows:

- **Home:** Clipboard, Fonts, Paragraph, Styles and Editing
- **Insert:** Pages, Tables, Illustrations, Links, Header & Footer, Text and Symbols
- **Page Layout:** Themes, Page Setup, Page Background, Paragraph and Arrange
- **References:** Table of Contents, Footnotes, Citations & Bibliography, Captions, Index and Table of Authorities
- **Mailings:** Create, Start Mail Merge, Write & Insert Fields, Preview Results and Finish
- **Review:** Proofing, Comments, Tracking, Changes, Compare and Protect
- **View:** Document Views, Show/Hide, Zoom, Window and Macros
- **Add-Ins:** PDF Transfer or any new add-in program

The Title Bar

The Title bar is next to the Quick Access toolbar. It displays the title of the current document which is in use. The first new document in Word is named as 'Document1' as shown below. When you open more new documents, Word automatically names them as Document2, Document3, Document4, etc., sequentially. The document can be saved by giving it a proper file name as per the user's choice.

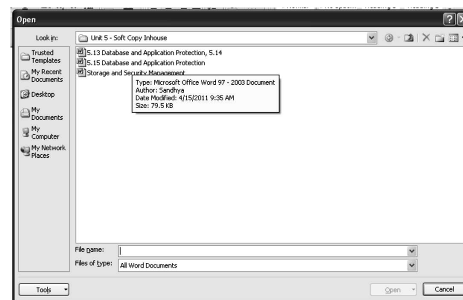


7.3.2 Opening Word Application

To open a Word 2007 application, you need to select the File Open menu as shown in the given screen and then select the file which you want to open.



After selecting the file the content of document file is opened on the screen.

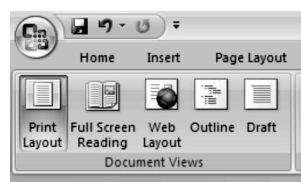


7.3.3 Explanation of Different Parts of Word Window and Views in Word

Following are the various ways in which a document can be viewed in Word:

- **Print Layout**: This is a view of the document as it would appear when printed. It includes all tables, text, graphics and images, as presented in the document.
- **Full Screen Reading**: This refers to a full-length view of a document. It is of utility when you want to view two pages at a time.
- **Web Layout**: This is a view of the document as it would appear in a Web browser.
- **Outline**: This is an outline of the document in the form of bullets.
- **Draft**: This view does not display pictures or layouts. It displays only the text.

To view a document in various forms, click on any of the document view shortcuts at the bottom of the screen or alternatively,



NOTES

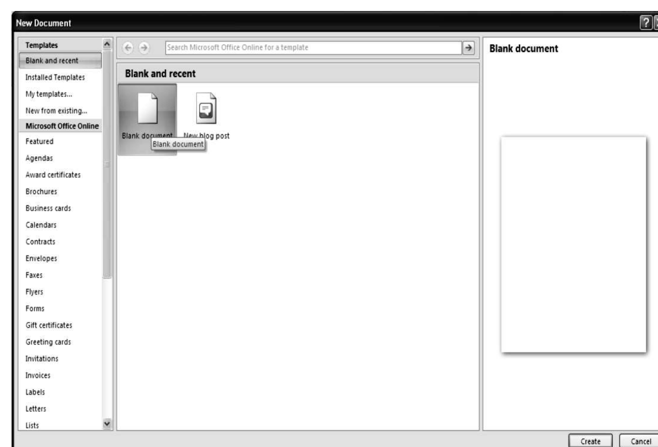
- Click on the **View Tab** in the Ribbon.
- Click on the appropriate document view.

A brief outline of the different types of views are as follows:

- ***Draft View*:** Draft view is the most frequently used view. The Draft view is used for the purpose of editing a document quickly.
- ***Web Layout*:** The Web Layout view enables you to see your document as it would appear in a browser such as Internet Explorer.
- ***Print Layout*:** The Print Layout view shows how the document will look when it is printed.
- ***Reading Layout*:** The Reading Layout view formats your screen so that you can can read the document more comfortably. You can choose more such reading options that appear on the right side of the user interface.
- ***Outline View***: The Outline view displays the document in an outline format. You can display headings without the text. If you move a heading, the accompanying text also moves with it.

7.3.4 Working with Documents

You can perform the following actions vis-à-vis documents:

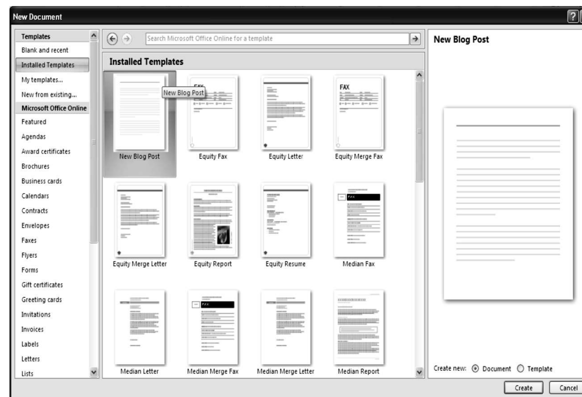


Creating a Document

Word provides various ways through which a user can create a new document, open an existing document and save a document in Word. To create a new document, Click on the **Microsoft Office Button** and then Click on **New** or Press **CTRL+N** on the keyboard.

You will see that when you click on the Microsoft Office Button and then Click on **New**, Word provides number of choices about the types of documents you can create. Select and click on **Blank** from the list if you want to create a blank document. You can create a document using the option **Installed Templates**.

Select any one of the templates as per your requirement. You can also browse other options through the list of choices that appear on the left.



Saving a Document

To save a document, the following steps need to be performed:

- Click on the **Microsoft Office Button** and Click on the **Save option**.
- The alternate option is to press **CTRL+S**.
- Click on the **File** icon in the Quick Access Toolbar.

Saving a Document using Save As

To save a document using the Save As option, Click on the **Microsoft Office Button** and Click on **Save As**. The Save As option helps you to save a document as a Word Document, Word Template, Word 97-2003 Document (earlier versions), Other Format, etc.

Renaming a Document

To rename an existing document, you need to perform the following steps:

- Click on the **Office Button** and locate the document you want to rename.
- Click on the **Save As** option and then right click on the document name with the mouse and select **Rename** from the shortcut menu.
- Type the new name for the document and press the **ENTER** key.

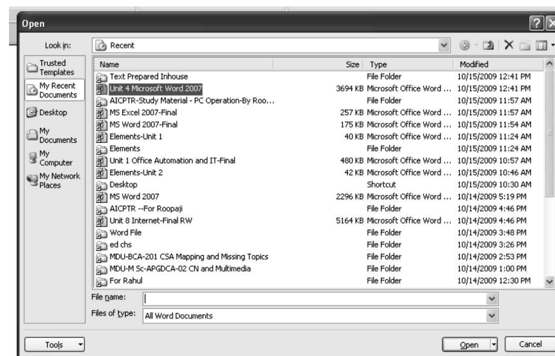
Working on Multiple Documents

Multiple documents can be simultaneously opened when you need to type or edit multiple documents at once. All the documents as opened will be listed in the **View Tab** of the Ribbon when you will click on Switch Windows. The current document has a checkmark beside the file name. You can select a different document as opened by simply clicking on the tab.

Opening an Existing Document

To open an existing document, the following steps need to be performed:

- Click on the **Microsoft Office Button** and then Click on **Open.**
- The alternate option is to press **CTRL+O** on the keyboard.
- For recently used document you can click on the **Microsoft Office Button** and then click on the name of the document in the **Recent Documents** window.



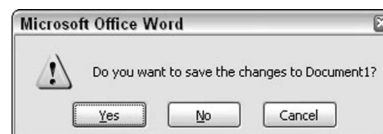
Closing a Document

To close a document, the following steps need to be performed:

- Click on the **Office Button.**
- Click on **Close.**

7.3.5 Quitting from Word

When you work with Word processing, you can either quit or minimize Word 2007. If do not expect to return to it anytime soon, you may just want to quit the program. If you want to stop work on one document to work on another, you can close the document and then open another. You can use the Minimize button to hide Word while you are off doing other things. Following steps are required to quit from Word 2007:



- Choose **Exit Word** from the Office Button menu.
- Save any files when Word prompts you to do so.
- Click **Yes** to save your file. You may be asked to give the file a name, if you have not yet done so.
- Click **Cancel** to 'quit' the Exit Word command and return to Word.

If you select quit, Word closes its Window and you return to Windows or some other program (if open) or to desktop. .

7.3.6 Text Editing

NOTES

The process of editing a document involves the following steps:

Typing and Inserting Text

To enter text, just type the text in the Word window. The text will appear at the location of the blinking cursor. You can move the cursor using the arrow keys on the keyboard or by positioning the mouse and clicking the left button. The keyboard shortcuts used for this purpose are as follows:

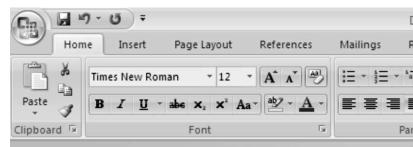
Move Action	*Keystroke*
Beginning of the line	HOME
End of the line	END
Top of the document	CTRL+HOME
End of the document	CTRL+END

To change the current attributes of the text as typed, it needs to be highlighted first. Select the text by dragging the mouse over the text to be modified while holding down the left mouse button. An alternate way is to hold down the **SHIFT** key on the keyboard and use the arrow buttons to highlight the text. Following are the shortcuts that are used to select a specific portion of the text:

Selection	*Technique*
Whole word	Double-click within the word
Whole paragraph	Triple-click within the paragraph
Several words or lines	Drag the mouse over the words or hold down the **SHIFT** key while using the arrow keys
Entire document	Choose **Editing** → **Select** → **Select All** from the Ribbon or simply press **CTRL+A**

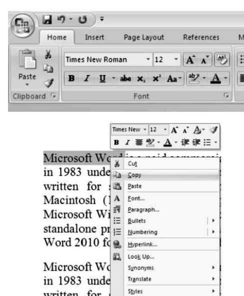
Moving and Copying Text

Moving and copying data are common commands used in many computer programs. These commands allow us to take information from one document or location and place them in another without retyping everything. When you move data, you are actually taking it from the location in which it is currently placed and relocating it to another area in the document as shown in the given screen.



NOTES

When you copy data, the original data remains intact and in addition a copy of that data is placed in another area in the document as shown in the given screen.



Using Drag and Drop Technique

Text can be inserted in a document at any point using any of the following methods:

Type Text: Put your cursor where you want to add/insert the text and begin typing.

Copy and Paste Text: Highlight the text you wish to copy and right click to view options. Now click on Copy option. Put your cursor where you want the text to be inserted in the document and right click to view options. Select Paste to paste the copied text.

Cut and Paste Text: Highlight the text you wish to cut and right click to view options. Now click on Cut option. Put your cursor where you want the text to be inserted in the document and right click to view the options. Select Paste to paste the cut text.

Drag Text: Highlight the text you wish to move. Click on it and drag it to the place where you want the text to be inserted in the document.

Using Cut, Copy and Paste Options from Edit Menu

To work with Cut, Copy and Paste operations from Edit menu in Word 2007, you need to first select the text which you want to copy and paste as shown in the given screen.

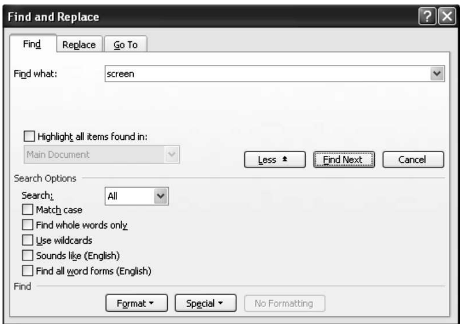


Select either Copy or Paste options or place the mouse pointer where you want to Paste the selected text. The copied and cut text will be stored in Clipboard application as shown in the following screen.



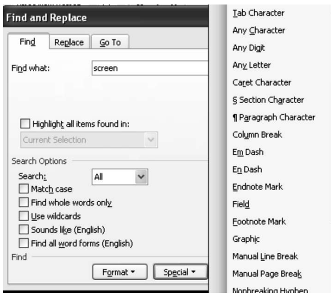
Finding and Replacing Text

The find and replace option can be accessed by selecting and pressing key combinations **CTRL+F** or **CTRL+H**. After choosing the Find or Replace option, you will get the shown screen.



The special drop down list drops a variety of options as follows:

The match case provides you to find and replace the word as uppercase or lowercase. For example, if you check on Match case box and type the word in capital as **'TOP'**, a dialog box appears with a message **'Word has finished searching the document. The search item was not found.'**



NOTES

If you remove the check box, the found word is replaced by defined word as shown:



You can also search the items by using wild card ('*') as shown below:



7.4 DOCUMENT PRINTING

Microsoft Word is considered as standardized software for creating text documents. In Word 2007, you can quickly print a document and can avoid the Print dialog box which saves your time. To use the Preview and Print functions in Microsoft Word 2007, follow the given steps.

- Open the document to preview, if you are not currently working on it.
- Click the **Microsoft Office button** in the top left hand corner of your screen. You will find numerous features here that used to be on the **'File'** menu in Microsoft Word 2007.
- Highlight **'Print'** and then select **'Print Preview'**. You will see the document on the screen.



Using the Print Function

In Microsoft Office button, three options, such as **Print**, **Quick Print** and **Print Preview** appear on the screen to perform the printing of the document.

Following steps are required to use a Print function in Microsoft Word 2007.

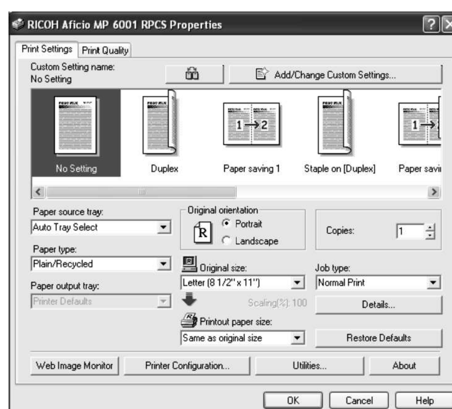
- Click the **Microsoft Office button** and click on **Print**.
- Select '**Quick Print**' if your document does not need any change.
- Choose '**Print**' if you want more printing options. The '**Print Dialog Box**' appears.
- Accept the default printer or choose a new one to the right of 'Name'.
- Click on the '**Properties**' or '**Options**' buttons to choose different options affecting your document's appearance.
- Select '**OK**' after you make your choices and the document will be printed.

7.4.1 Setup Page to Margin, Paper Size, Orientation for Printing

To print the document you need to check the options on the Ribbon. It tell you that now you are in the 'Preview' mode. You can change margin size, page orientation, print options and can select many other print functions in the 'Preview' mode as shown in the given screen.

NOTES

Select Properties button if you want to view the margin, paper size and orientation to print the selected Word document.

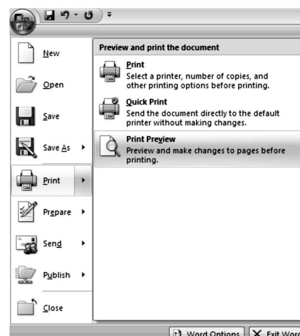


7.4.2 Print Preview

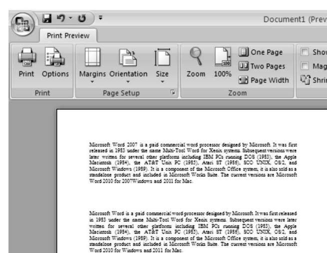
Print Preview option is used to see how the document will look like when it is printed on paper. You can get into Print Preview mode by clicking the **Print Preview** icon on the Quick Access Toolbar or clicking on the **Office Start Menu** → **Print** → **Print Preview**.



You can also select the Print Preview option in Word 2007 from Office button as shown in the given screen.



After clicking on **Print Preview** option form **Print** menu, you can see the previewed pages and can modify the required changes.



NOTES

7.5 DOCUMENT FORMATTING

Document formatting represents all the types of formatting part which are applied to a selected document for making suitable presentation of the Word Document. The page margins can be modified through the following steps:

- Click on the **Page Layout** Tab on the Ribbon.
- On the **Page Setup** Group, Click on **Margins.**
- Click on **Default Margin.**
- Or you can also click on **Custom Margins** and complete the dialog box.

Following steps are required to change the Orientation, Size of the Page or Columns:

- Click the **Page Layout** Tab on the Ribbon.
- On the **Page Setup** Group, Click on the **Orientation**, **Size** or **Columns** drop down menus.
- Click on the appropriate choice.

To Apply a Page Border and Color

To apply a page border or color following steps are required:

- Click on the **Page Layout** Tab on the Ribbon.
- On the **Page Background** Group, click on the **Page Colors** or **Page Borders** drop down menus.

NOTES

To Insert a Page Break

To insert a page break, following steps are required:

- Click on the **Page Layout** Tab on the Ribbon.
- On the **Page Setup** Group, click on the **Breaks Drop Down Menu.**
- Click on **Page Break.**
- Select the option as per requirement.



To Insert a Cover Page

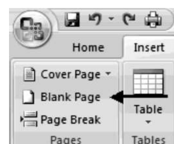
To insert a cover page, following steps are required:

- Click on the **Insert** Tab on the Ribbon.
- Click on the **Cover Page** option on the Page Group.
- Choose a style for the cover page.

To Insert a Blank Page

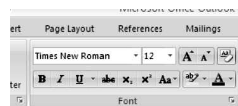
To insert a blank page, following steps are required:

- Click on the **Insert** Tab on the Ribbon.
- Click on the **Blank Page** option on the Page Group.



7.5.1 Font Formatting

To change the font typeface, click on the arrow next to the font name and select a font.





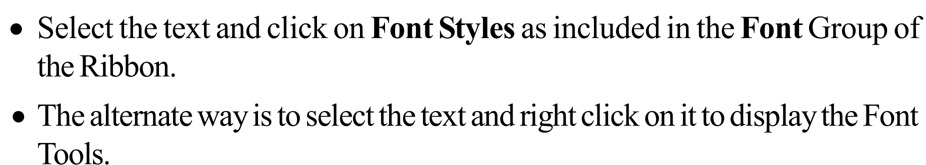
NOTES

You can preview how the new font will look by highlighting the text and then hovering over the new font typeface.

To change the font size, click on the arrow next to the font size and select the proper size or click on the increase or decrease font size buttons.

Font Styles and Effects

Font styles are predefined formatting options which are used to emphasize specific text. They include bold, italics and underline. Following are the steps that need to be followed in order to add these effects:



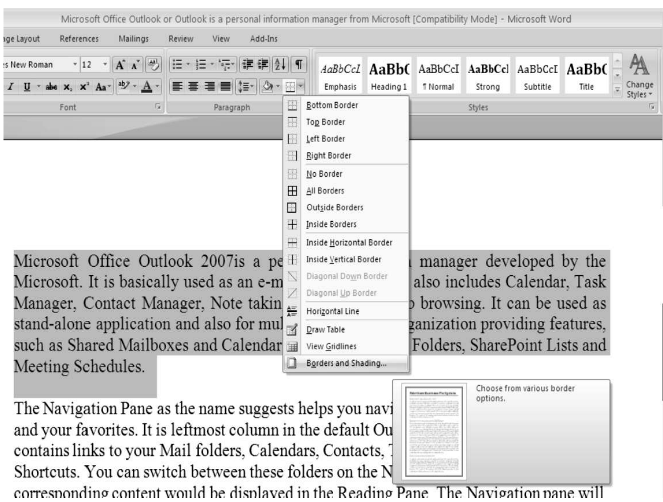
- Select the text and click on **Font Styles** as included in the **Font** Group of the Ribbon.
- The alternate way is to select the text and right click on it to display the Font Tools.

7.5.2 Border and Shading

You can add borders, shade paragraphs and even entire pages. To create a border around a single paragraph or paragraphs, you need to do the following:

- Select the area of text where you want to add the border or to add shade.
- Click on the **borders** Button in the **Paragraph** Group on the **Home** Tab.
- Select the **Border and Shading** option.
- Select the options as per your choice and requirement.

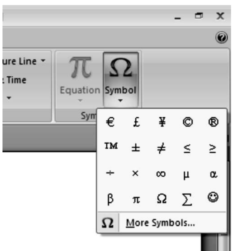
NOTES



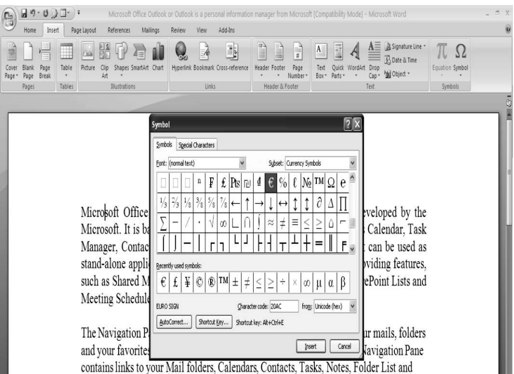
7.5.3 Inserting Symbols

Word 2007 permits you to insert special characters, symbols, pictures, illustrations, etc. Special characters are punctuation, spacing or typographical characters that are generally not available on the standard keyboard. Following are the steps for inserting symbols and special characters:

- Move your cursor in the document where you want to insert the symbol.
- Click on the **Insert** Tab in the Ribbon.
- Click on the **Symbol** button in the **Symbols** Group.
- Select the symbol as required.



If you want more symbols then click on **More Symbols** to display the following dialog box for a list of various symbols in various fonts.

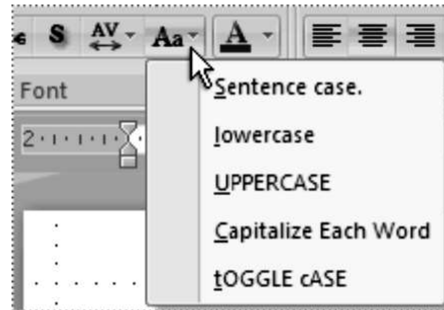


NOTES

7.5.4 Change Case

In Microsoft Word 2007, you can change the capitalization of words, sentences or paragraphs by doing the following:

- Select the text for which you want to change the case.

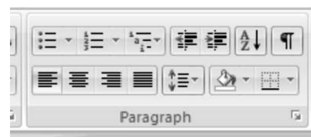


- On the **Home** tab, in the **Font** group, click on **Change Case** Aa (as shown in the screen) and then click on the capitalization option that you want.
- To capitalize the first letter of a sentence and leave all other letters as lowercase, click on **Sentence case**.
- To exclude capital letters from your text, click on **lowercase**.
- To capitalize all of the letters, click on **UPPERCASE**.
- To capitalize the first letter of each word and leave the other letters lowercase, click on **Capitalize Each Word**.
- To shift between two case views, for example, to shift between **Capitalize Each Word** and the opposite, **cAPITALIZE eACH wORD**), click on **tOGGLE cASE**.

7.5.5 Text Alignment

Paragraph alignment helps in organizing the text according to the way you want it to appear. To change the alignment, follow the given steps:

- Click on the **Home Tab.**
- Select the appropriate button for alignment in the Paragraph Group.



  - o **Align Left:** The text is aligned with the left margin.
  - o **Center:** The text is centered within the margins.
  - o **Align Right:** The text is aligned with the right margin.
  - o **Justify:** It aligns the text to both the left and right margins.

7.5.6 Insertion of Current Date and Time in a Word Document

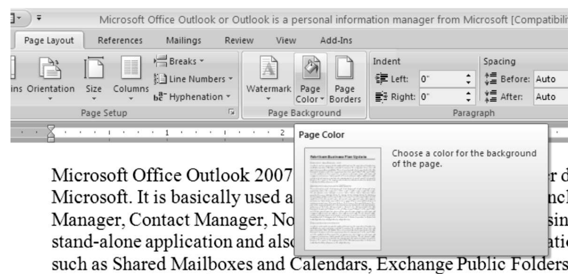
In Microsoft Word 2007, use the following shortcut keys to enter the current system date and time as shown in the given screen, for example [**Alt+Shift+D**] key combination is used to insert the current date and [**Alt+Shift+T**] key combination is used to insert the current time.

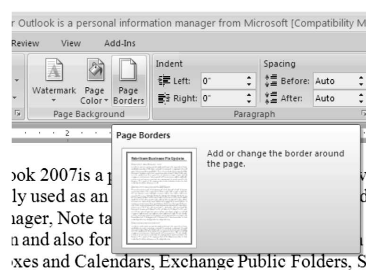
> Microsoft Word 2007 is a paid commercial word processor designed released in 1983 under the name Multi-Tool Word for Xenix systems. later written for several other platforms including IBM PCs runnin Macintosh (1984), the AT&T Unix PC (1985), Atari ST (1986), Microsoft Windows (1989). It is a component of the Microsoft Office standalone product and included in Microsoft Works Suite. The curr Word 2010 for 2007Windows and 2011 for Mac.
>
> Current Date: 4/15/2011 Time: A10:58 AM

7.5.7 Writing Header/Footer Content

To apply a page border or color, follow the given steps:

- Click on the **Page Layout** Tab in the Ribbon.
- In the **Page Background** Group, click on **Page Color** or **Page Borders** to generate the drop down menus.

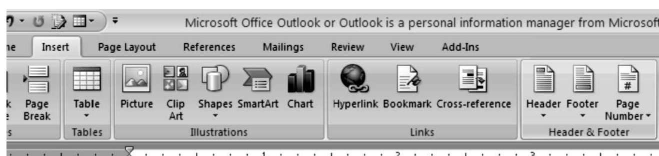




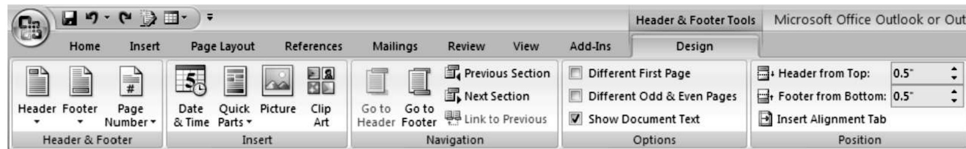
Inserting Common Header and Footer Information

To insert Header and Footer information, such as page numbers, date or title, first decide whether you want the information to be in the Header, i.e., at the top of the page or in the Footer, i.e., at the bottom of the page and then follow the given steps:

- Click on the **Insert** Tab in the Ribbon.
- Click on **Header & Footer.**
- Select a style.



- The **Header & Footer Design** Tab will be displayed in the Ribbon.
- Select or type the information that you want to have in the Header or Footer, such as date, time, page numbers, etc.



7.5.8 Bullets and Numbering

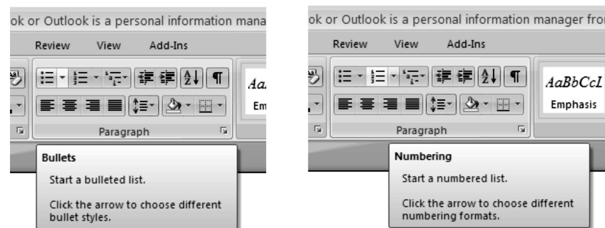
Bulleted lists are represented using bullet points, numbered lists are represented using numbers and outline lists combine together the numbers and letters depending on the association of the list.

Following are the required steps for adding a list to existing text:

- Select the text for which you wish to make a bulleted or numbered list.
- From the Paragraph Group on the Home Tab, click on the **Bulleted** or **Numbered Lists** button as per requirement.

To **create** a new list:

- Move your cursor in the document where you want the list to appear.
- Click on the **Bulleted** or **Numbered Lists** button.
- Start typing.
- Bullets or Numbers will automatically appear when you will press the **ENTER** key and go to a new line.



Nested Lists

A nested list is the list which has several levels of indented text. To create a nested list, do the following steps:

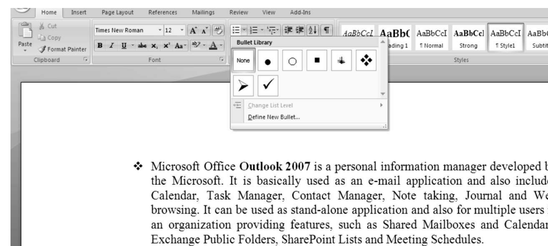
- Create your list following the directions as given above.
- Click on the **Increase or Decrease Indent** button.



Formatting Lists

The bullet image and the numbering format can be modified using the **Bullets** or **Numbering** dialog box. Following are the steps for formatting a bulleted or a numbered list:

- Select the entire list to change all the bullets or numbers or just place the cursor on one line within the list to change a single bullet.
- Right click to view options.
- Click on the arrow next to the bulleted or numbered list and choose the bullet or numbering style as required.



7.6 ADVANCED DOCUMENT FORMATTING

Styles include collection of formatting that you can apply to text. To utilize **Quick Styles** follow the given steps**:**



- Select the text you wish to format.
- Click on the dialog box next to the **Styles Group** on the Home Tab.
- Click on the style you wish to apply.

The use of Styles in Word 2007 will allow you to quickly format a document with a consistent and professional look. Styles can be saved for use in other documents.

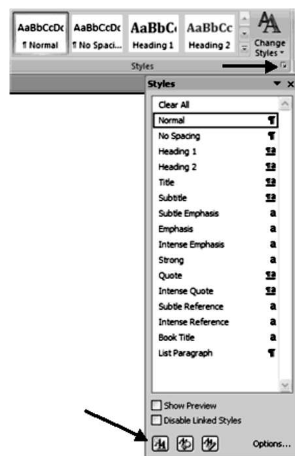


Apply Existing Styles

There are many styles that are already in Word ready for you to use. To view the available styles click on the Styles dialog box on the **Styles Group** in the **Home Tab**. To apply a style follow the given steps:

- Select the text.
- Click on the Styles Dialog Box.
- Click on the Style as per your requirement.

NOTES



Creating New Styles

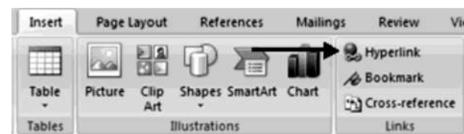
You can create styles for formatting that you use regularly. There are two ways to do this which are known as **New Styles** or **New Quick Styles**.

Creating Hyperlinks

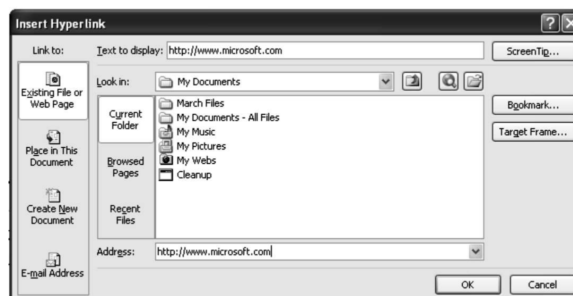
A hyperlink or link is a reference to data that the reader can directly follow or that is followed automatically. A hyperlink points to a whole document or to a specific element within a document. Hypertext is text with hyperlinks. A user following hyperlinks is said to navigate or browse the hypertext. A hyperlink has an anchor, which is the location within a document from which the hyperlink can be followed. The document containing a hyperlink is known as its source document.

Creating links or hyperlinks in a Word 2007 document allows you to put in a Uniform Resource Locator (URL) which helps the readers to directly click on to visit a Web page. To insert a link follow the given steps:

- Click on the **Hyperlink Button** on the **Links** Group of the Insert Tab.
- Type in the text in the 'Text to Display' box and the Web address in the 'Address' box.



Once you click on Hyperlink tool, following screen will appear.



NOTES

7.6.1 Indention

Indention helps you to set the text within a paragraph at different margins. The various options for indenting text are as follows:

- **First Line:** Controls the left boundary for the first line of a paragraph.
- **Hanging:** Controls the left boundary for every line in a paragraph except the first one.
- **Left:** Controls the left boundary for every line in a paragraph.
- **Right:** Controls the right boundary for every line in a paragraph.

Following are the steps that need to be performed for indenting paragraphs:

- Click on the **Indent** buttons to control the indent.
- Click on the **Indent** button repeatedly to increase the size of the indent.
- Click on the dialog box of the **Paragraph** Group.
- Click on the **Indents and Spacing** Tab.
- Select the indent as required.



7.6.2 Using Tab

Tab stops are frequently used to create easy-to-format documents. You can use the ruler to set manual tab stops at the left side, middle and right side of your document. If the horizontal ruler that runs along the top of the document is not visible then to view it click on the **View Ruler** button at the top of the vertical scroll bar. You can quickly set tabs by clicking on the tab selector at the left end of the ruler until it displays the type of tab that you require and then clicking the ruler at the location you require. If you want your tab stops at specific positions that you can not get by clicking the ruler or if you want to insert a specific character before the tab, then you can use the **Tabs** dialog box. Following are the various Tab tools used for Tab setting and their functions.

NOTES

Tools Used For Tab Setting	*Function*
⌞	A **Left Tab** stop sets the start position of text that will then run to the right as you type.
⊥	A **Center Tab** stop sets the position of the middle of the text. The text centers on this position as you type.
⌟	A **Right Tab** stop sets the right end of the text. As you type, the text moves to the left.
⊥.	A **Decimal Tab** stop aligns numbers around a decimal point. Independent of the number of digits, the decimal point will be in the same position. You can align numbers around a decimal character only and you cannot use the decimal tab to align numbers around a different character, such as a hyphen or an ampersand symbol.
I	A **Bar Tab** stop does not position text. It only inserts a vertical bar at the tab position.

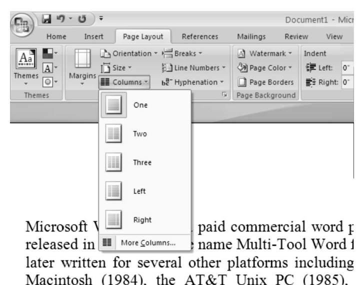
7.6.3 Dividing Page of a Document into Multiple Columns

Microsoft Word 2007 can be used for writing term papers, articles or even business reports. It can also be used to create newsletters, flyers or stationery. When you become acquainted with the various functions of Microsoft Word, you will find that you can have a lot of fun using it to create many different things. In order to create something like a newsletter or a shopping list, you may find it the best option as it divides your document into several columns. This is not hard to do and it looks rather impressive and professional. Following steps are required to divide page of a document into multiple columns:

- Open the document in Microsoft Word 2007.
- Select the text you want to divide into columns before proceeding. If you have not typed any text, follow these steps and the columns will appear as you type.
- Click on the **'Page Layout'** menu. In the Page Setup section, you will see a button labeled 'Columns'.
- Click on the **'Columns'** button. You will see several choices for columns click on the one you want. If you want to customize the columns, click on **'More Columns'** at the bottom.
- Select the option for how many columns you want and how wide to make each one. The preview box will show you how it will look like.

NOTES

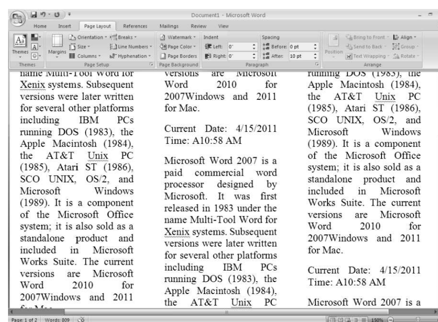
- Click on **'OK'**. Your document will be formatted in columns as shown in the given screen.



Select the given option from **More Columns…** You will get the screen where you can select the multi columns for documents as shown in the given screen.

7.6.4 Inserting Contents from another File into the Current Document

You can insert contents from another file, such as linking to a location in a document or Web page that you created in Word 2007. You must mark the hyperlink location or destination and then add the link to it. A hyperlink is a graphic or a piece of text that links to another Web page. To mark the hyperlink location, following steps are required:



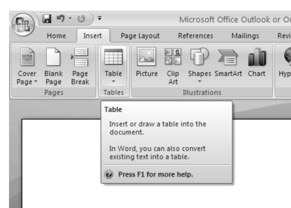
- Insert a bookmark in the destination file or Web page.
- Open the file that you want to link from and select the text or object you want to display as the hyperlink.
- Right click and then click on **Hyperlink** on the shortcut menu.
- Under **Link to**, click on **Existing File or Web Page**.
- In the **Look in** box, click on the down arrow. Then navigate to and select the file that you want to link to.
- Click on **Bookmark**, select the bookmark that you want and then click on **OK**.

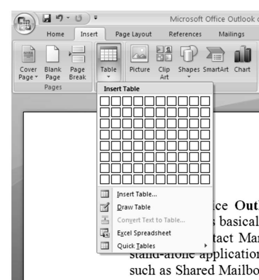
To customize the **ScreenTip** that appears when you rest the pointer over the hyperlink, click on **ScreenTip** and then type the text that you want to insert. If you do not specify a tip, Word 2007 uses the path to the file including the bookmark name as the tip.

NOTES

7.7 CREATING AND USING TABLE

Tables organize data into rows and columns. They are used for the purpose of displaying data in a tabular format.





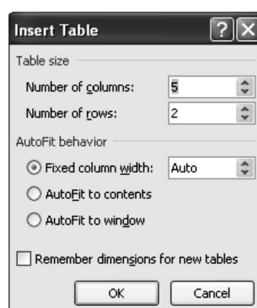
To Create a Table

To create a table, follow the given steps:

- Move the cursor on the page where you want the new table.
- Click on the **Insert** Tab of the Ribbon.
- Click on the **Tables** Button in the Tables Group.

Following are the four ways in which tables can be created:

- o Highlight the number of rows and columns.
- o Click on **Insert Table** and enter the number of rows and columns.
- o Click on **Draw Table**, create your table by clicking and entering the rows and columns.
- o Click on **Quick Tables** and select a table.

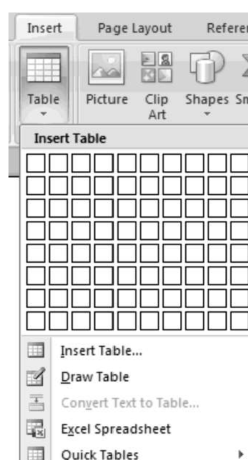


To enter data in a table, place the cursor in the cell where you want to enter the data or information. Start typing to enter the data.

7.7.1 Inserting a Table into the Document

To insert a table into the document, you need to select the number for rows and columns. After selecting the required rows and columns from Insert Table tab, you can get a table format in which you can work.

NOTES



- Place the cursor on the page where you want the new table.
- Click on the **Insert** Tab of the Ribbon.

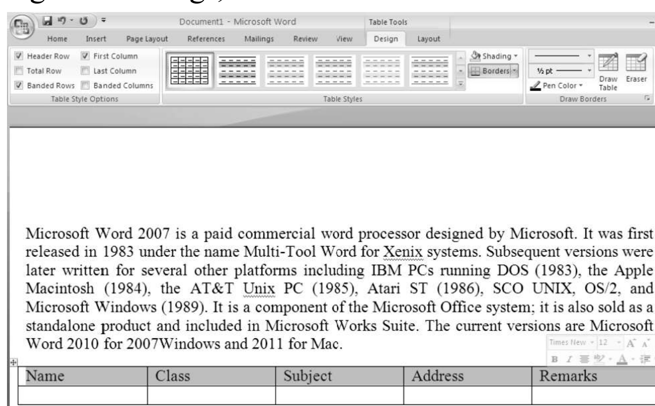
 Click on the **Tables** Button on the Tables Group.

You can create a table using one of four following ways:

- Highlight the number of row and columns.
- Click on **Insert Table** and enter the number of rows and columns.
- Click on the **Draw Table**, create your table by clicking and entering the rows and columns.
- Click on **Quick Tables** and choose a table.

7.7.2 Selection within Table

Selection within tables is prime operation of Word 2007 because it helps you to select and perform the desired options, such as finding total marks of a specific student or merge the headings, etc.



Microsoft Word 2007 is a paid commercial word processor designed by Microsoft. It was first released in 1983 under the name Multi-Tool Word for Xenix systems. Subsequent versions were later written for several other platforms including IBM PCs running DOS (1983), the Apple Macintosh (1984), the AT&T Unix PC (1985), Atari ST (1986), SCO UNIX, OS/2, and Microsoft Windows (1989). It is a component of the Microsoft Office system; it is also sold as a standalone product and included in Microsoft Works Suite. The current versions are Microsoft Word 2010 for 2007Windows and 2011 for Mac.

Name	Class	Subject	Address	Remarks

You can also set the various options for selected table as shown in the following screen.

Name	Class	Subject	Address	Remarks

7.7.3 Insert/Delete Row(s)/Column(s) into Table

To insert rows and columns, you need to select shortcut keys which provide options, such as Insert Columns to the Left, Insert Columns to the Right, Insert Rows Above and Insert Rows Below. After selecting the option, for example Insert Columns to the Right, you will get an extra column to the existing eight most column.

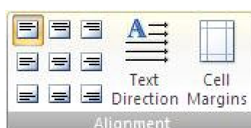
7.7.4 Applying Borders and Shading in the Table

Borders and shading facilitate you to highlight information in a table. Thus it enhances the appearance of the text in a table. Word automatically applies a ½ point border and a white background, i.e., no shading to all tables and table cells. Though, to emphasize definite portions of your table, you can add, remove or modify table borders or you can add shading to certain cells, rows or columns in your table. In Word 2007, the Ribbon provides quick access to the table borders and shading features. You can add borders to a table or individual table cell as shown.

Competitor Ranking	Current Share	Share in 3 Years
Largest Competitor	40%	25%
Second Largest Competitor	30%	18%
Third Largest Competitor	20%	15%

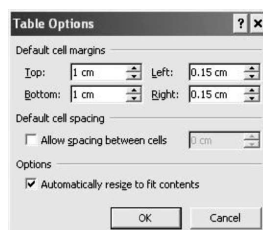
7.7.5 Formatting Data within Table Cells

The Ribbon provides options for changing both the vertical and horizontal alignment with cells.



- Select the cells you want to format.
- Click on the **Layout** tab under Table Tools on the Ribbon.
- Click on on the required button from the Alignment group.

The cell margins allow you to change the gap between the cell contents and the border of the cell.



NOTES

- Select the cells you want to format or the entire table.
- Click on the **Layout** tab on the Ribbon.
- Click on the **Cell Margins** button in the Alignment group.
- The **Table** Options dialog box appears. Adjust the required margin settings.

Adjust the measurement in the **Allow spacing between cells** box to change the spacing between the cells as opposed to within the cell between the content and its borders. Click on **OK.** Your settings become the default settings for all future tables.

7.7.6 Table Autoformatting

By selecting certain formatting options, the Microsoft Office 2007 system programs can automatically format certain kinds of text as you type, such as replacing a typed hyphen (-) with a dash (–). Use the AutoFormat As You Type options to control the automatic formatting that you want. The specific **AutoFormat As You Type** options that are available depend on the program that you are using. There are two ways to use the capabilities of the AutoFormat feature:

- You can use features in Microsoft Office Word and Microsoft Office Outlook. The **AutoFormat As You Type** tab includes the most useful capabilities of the AutoFormat feature.
- In Word, you can add the AutoFormat command to the Quick Access Toolbar.

7.8 USING TOOLS

This step merges the variable information with the text or the letter. You can output the merge result by using either of the following options:

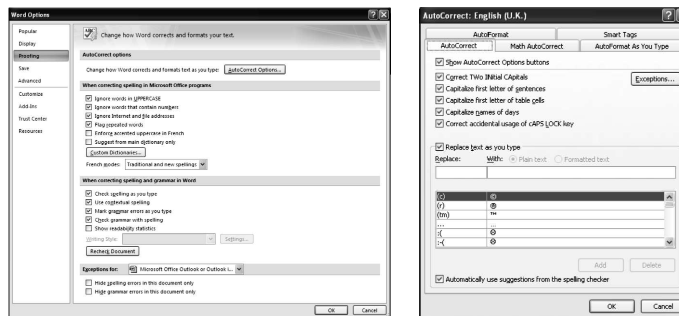
- **Print**: Select this option to send the merged document directly to the printer. You will not be able to view the document on your screen. When you click on **Print**, the **Merge to Printer** dialog box appears. In the **Merge to Printer** dialog box, you can choose which records to merge. When you click on **OK,** the Print dialog box appears. Click **Print** to print the merge document.
- **Edit Individual Letters**: Select this option to display the merged document on your screen. When you click on **Edit individual letters**, the **Merge to New Document** dialog box appears. In the **Merge to New Document** dialog box, you can choose which records to merge. When you click on **OK**, the documents are merged to a new Word document. To print the file, on the **File** menu, click on **Print**.

In Word 2007, click on the **Microsoft Office Button** and then click on **Print**.

7.8.1 AutoCorrect

AutoCorrect option is used to correct the common typing or spelling errors. You can set up the AutoCorrect tool in Word 2007 to keep some text as it is. The following steps have to be followed to customize AutoCorrect:

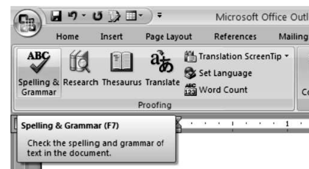
- Click on the **Microsoft Office** button.
- Click on the **Word Options** Button.
- Click on the **Proofing** tab.
- Click on the **AutoCorrect Options** button.



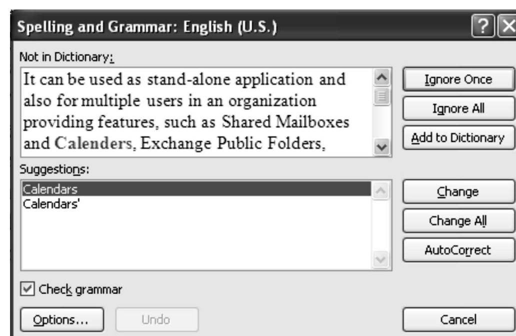
- In the **AutoCorrect Tab**, specify the words which you want to replace whenever you type.

7.8.2 Spell Checking

To check the spelling and grammar of a document, follow the given steps:



- Move the cursor to the beginning of the document or to the beginning of the section that you want to check.
- Click on the **Review** Tab in the Ribbon.
- Click on the **Spelling & Grammar** option in the Proofing Group.
- Any error as encountered will be displayed in a dialog box which would permit you to select a more proper spelling or phrasing as shown below.



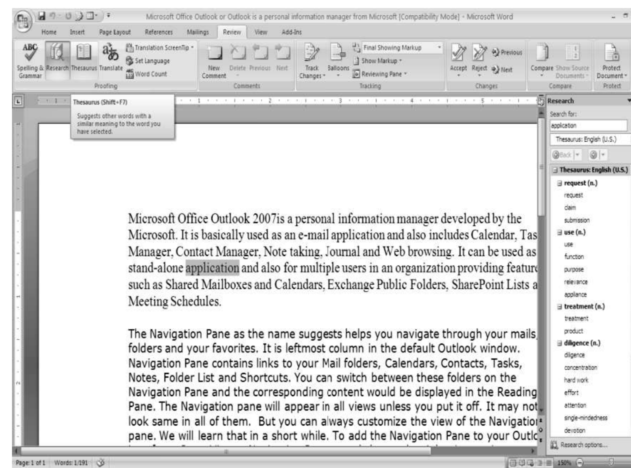
If you wish to check the spelling of an individual word, you can right click on any word that has been underlined by Word and select a proper substitution.

Thesaurus

NOTES

Thesaurus allows you to view synonyms. To use the thesaurus follow the given steps:

- Click on the **Review** Tab of the Ribbon.
- Click on the **Thesaurus** Button on the Proofing Group.
- The thesaurus tool appears on the right side of the screen and you can view the word options.



7.8.3 Protecting File

Following steps are required to protect a file:

- Click on the **Microsoft Office Button** and then click **Save As option.**
- Click on **Tools** and then click on **General Options**.
- You can type a password in **Password to open** or **Password to modify**.
- Click on **OK**.
- When prompted, retype your passwords to confirm them and click on **OK**.
- Click on **Save**.

7.9 INSERTING GRAPHICS IN WORD DOCUMENT

In Word 2007, you can modify objects and include graphics in many ways to add sparkle to even the most mundane document. You can add drop caps, group two or more objects into one object and change the layering of overlapping objects. You can also position graphics anywhere on a page, change text wraps around an object and modify how objects are aligned relative to each other and to the page.

NOTES

You can enhance graphic objects with gradient color fills, textures, shadows and 3-D effects and by adding a watermark and page border.

Word 2007 allows you to insert illustrations and pictures into a document. To insert illustration do the following:



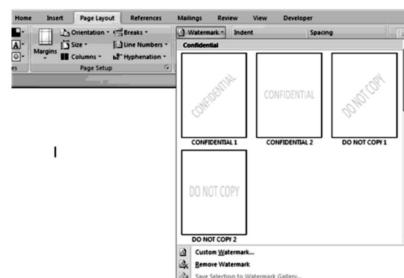
- Place your cursor in the document where you want to insert the illustration/ picture.
- Click on the **Insert** Tab on the Ribbon.
- Click on the **Clip Art** button.
- The dialog box will open on the screen and you can search for Clip Art.
- Choose the illustration you wish to include.

To insert a picture, following steps are required:

- Place your cursor in the document where you want to insert the illustration/ picture.
- Click on the **Insert** Tab on the Ribbon.
- Click on the **Picture** button.
- Browse to the picture you wish to include.
- Click on the **Picture.**
- Click on **Insert.**

Watermark is a translucent image that appears behind the primary text in a document. To insert a watermark do the following:

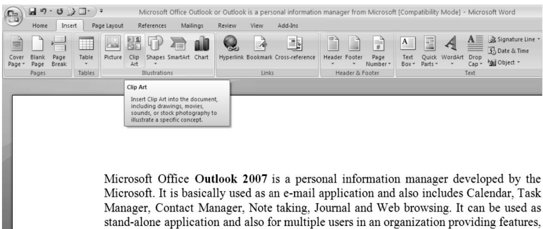
- Click on the **Page Lay**out Tab in the Ribbon.
- Click on the **Watermark** button in the **Page Background** Group.
- Click on the **Watermark** you want for the document or click on **Custom Watermark** and create your own watermark.
- To remove a watermark, follow the steps above but click on **Remove Watermark.**



NOTES

7.9.1 Inserting Clip Art, WordArt, AutoShapes and Text Box

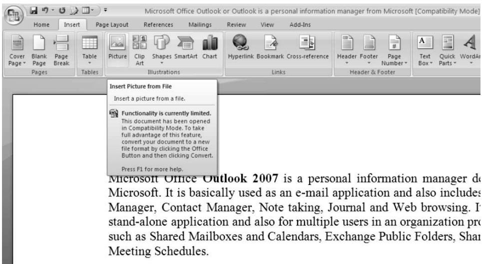
Word 2007 permits you to insert illustrations and pictures into the document, at the desired location. Following are the steps for the insertion of **illustrations** into a document:

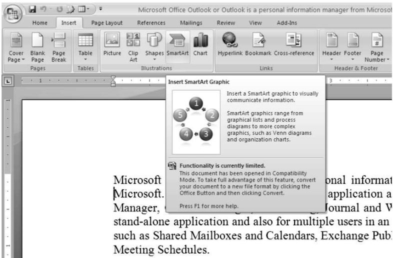


- Move your cursor in the document where you want the illustration/picture to appear.
- Click on the **Insert** Tab on the Ribbon.
- Click on the **Clip Art** Button.
- The dialog box will open on the screen and you can search for Clip Art.
- Select the illustration that you want to use in the document.

Smart Art is a collection of graphics that you can utilize to organize information in your document. It includes timelines, processes or work flow. The following steps need to be performed in order to insert SmartArt ino the document.

- Move your cursor in the document where you want the illustration/picture to appear.
- Click on the **Insert** Tab in the Ribbon.
- Click on the **SmartArt** button.
- Click on the **SmartArt** that you wish to include in your document.
- Click on the arrow on the left side of the graphic to insert text or type the text in the graphic.





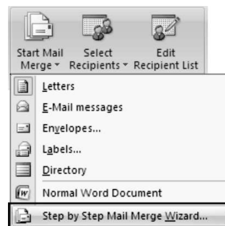
7.10 MAIL MERGE

NOTES

Mail Merge is one of the effective features of Microsoft Word. It helps you to draft and send letters to many recipients in very little time. You can use mail merge when you want to create a set of documents, such as a form letter that is sent to many customers. Each letter has the same kind of information yet the content is unique. For example, in letters to your customers, each letter can be personalized to address each customer by name. The unique information in each letter comes from entries in a data source. The mail merge process entails the following overall steps:

- **Set up the Main Document:** The main document contains the text and graphics that are the same for each version of the merged document, for example the return address or salutation in a form letter.
- **Connect the Document to A Data Source:** A data source is a file that contains the information to be merged into a document, for example the names and addresses of the recipients of a letter.
- **Refine the List of Recipients or Items:** Microsoft Office Word 2007 generates a copy of the main document for each item or record in your data file. If your data file is a mailing list, these items are probably recipients of your mailing. If you want to generate copies for only certain items in your data file, you can choose which items (records) to include.

Add placeholders, called mail merge fields, to the document. When you perform the mail merge, the mail merge fields are filled with information from your data file. Preview and complete the merge. You can preview each copy of the document before you print the whole set. You use commands on the **Mailings** tab to perform a Mail Merge. If you plan to use custom contact fields with the Contacts list in Microsoft Office Outlook, you must begin the mail merge process in Outlook. You can also perform a Mail Merge by using the Mail Merge task pane which leads you step by step through the process. To use the task pane, in the Start Mail Merge group on the **Mailings** tab, click **Start Mail Merge** and then click on **Step by Step Mail Merge Wizard**.



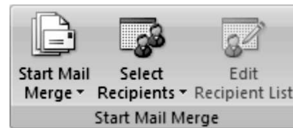
7.10.1 Creating a Master Document

Following steps are required to create the main document:

- Start Microsoft Word 2007. A blank document opens by default. Leave it open. If you close it, the commands in the next step are not available. On

NOTES

the **Mailings** tab, in the **Start Mail Merge** group, click on **Start Mail Merge**.

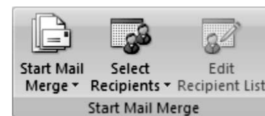


Click on Letters. You can also use Mail Merge to create the following:

- You can create a catalog or directory which has the same kind of information, such as name and description is shown for each item but the name and description in each item is unique. Click on Directory to create this type of document.
- You can create a set of envelopes. The return address is the same on all the envelopes, but the destination address is unique on each one.
- You can create a set of mailing labels. Each label shows a person's name and address but the name and address on each label is unique.
- You can create a set of e-mail messages. The basic content is the same in all the messages but each message goes to the individual recipient and each message contains information that is specific to that recipient, such as the recipient's name or some other piece of information.

7.10.2 Creating Data Source

To merge information into your main document, you must connect the document to a data source or a data file. If you do not already have a data file, you can create one during the Mail Merge process. If you use an existing list, make sure that it contains the information that you want to use, including all the columns and the rows. You can make some changes during the merge but you cannot open your data source separately during the merge. The merge process is easier if your data source is ready before you connect to it. Choose a data file. On the **Mailings** tab, in the **Start Mail Merge** group click on **Select Recipients**.



Do one of the following tasks:

- If you want to use your Contacts list in Outlook, click on **Select from Outlook Express Contacts**.
- If you have a Microsoft Office Excel worksheet, a Microsoft Office Access database or another type of data file, click on **Use Existing List** and then locate the file in the **Select Data Source** dialog box.

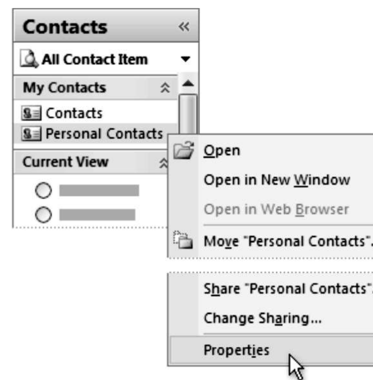
NOTES

For Excel, you can select data from any worksheet or named range within a workbook. For Access, you can select data from any table or query. The query includes a means of finding all the records stored in a data source that fit a set of criteria that you name. Queries can contain operators, quotation marks, wildcard characters and parentheses to help focus your search, i.e., defined in the database. For another type of data file, select the file in the Select Data Source dialog box. If the file is not listed, select the appropriate file type or select All Files in the Files of type box. In a Mail Merge, you can use the following types of data files:

- Files from single tier supports file based database programs for which you have installed an OLE DB. In fact, OLE DB is a component database architecture that implements efficient network and the Internet access to many types of data sources, including relational data, mail files, flat files and spreadsheets provider or Open Database Connectivity (ODBC).
- A standard method for sharing data between databases and programs. ODBC drivers use the standard Structured Query Language (SQL) to gain access to external data.

The first row of the table must contain column names and the other rows must contain data. Any similar address lists that have been created with a Messaging Application Programming Interface (MAPI) allows different messaging and workgroup applications including e-mail, voice mail and fax to work through a single client, such as compatible messaging system via as Microsoft Outlook or Microsoft Word document. You can merge tables into a single table as per your requirement. The first row of the table must contain headings and the other rows must contain the records that you want to merge. You can also use a header source. A document that contains the header row or header record to be used with the data source specified for a Mail Merge main document as a data source. Data fields represent a category of information that corresponds to one column of information in a data source. The name of each data field is listed in the first row (header row) of the data source, such as ‘PostalCode’ and ‘LastName’ are examples of data field names which are separated or delimited by tab characters or commas and data records. A complete set of related information that corresponds to one row of information in the data source represents data records. All information about one client in a client mailing list is an example of a data record which is separated by paragraph marks. Create a new data file in Word 2007. If you do not have a data file then click on **Type New List** and then use the form that opens to create your list. The list is saved as a database (.mdb) file that you can reuse, for example a Microsoft Excel list or a Microsoft Access database. If you are ready to continue, go to the next step. In Microsoft Office Outlook, on the Go menu click on Contacts. Right click on the Contacts folder that contains the information that you want to use for a Mail Merge and then click on Properties.

NOTES



An Outlook user profile is a group of e-mail accounts and address books. Usually, you need only one profile but if you share your computer with other people or use different address books for different purposes, you can set up more than one profile. When you are prompted for a profile, choose the profile that includes the Contacts folder that you want to use in the Mail Merge.

To view a profile, do the following:

- Click on **Show Profiles**.
- To be prompted to select a profile each time you start Outlook, click on Prompt for a profile to be used and then click on **OK**.

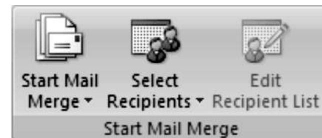
If your data file is an Excel worksheet that includes percentages, currency values or postal codes you can preserve the numeric formatting of the data by using Dynamic Data Exchange (DDE) which will connect the Excel worksheet from Word 2007. For example, you can make sure a five digit postal code of 07865 from your data file is not displayed as the number 7865 (without the leading zero). Before you connect to the worksheet, do the following in Word 2007:

- Click on the **Microsoft Office Button** and then click on **Word Options**.
- Click on **Advanced**.
- Scroll to the General section and select the **Confirm file format conversion** on open check box.
- Click on **OK**.
- With the Mail Merge main document open, in the **Start Mail Merge** group of the **Mailings** tab, click on **Select Recipients** and then click on **Use Existing List**.
- Locate the Excel worksheet in the **Select Data Source** dialog box and double click on it.
- In the **Confirm Data Source** dialog box, click on Microsoft Excel Worksheets via DDE (*.xls) and then click on **OK**.

In the **New Address List** dialog box, type the information for the first address or record that you want to include in your label merge. If you want to add or remove columns, click on **Customize Columns** and then make the changes

NOTES

that you want. After you type all the information for your first record, click on **New Entry** and then type the information for the next record. Continue until you have typed information for all the records that you want to include. When you connect to a certain data file, you might not want to merge information from all the records in that data file into your main document. If you plan to use your entire list, you can skip ahead to the next step. To arrange the list of recipients or use a subset of the items in your data file, do the following:



- On the **Mailings** tab, in the Start **Mail Merge** group, click on **Edit Recipient List**.

In the Mail Merge Recipients dialog box you need to select the names of recipients to whom you want to send the mail.

To filter records, do the following:

- Under Refine recipient list, click on **Filter**.
- On the Filter Records tab of the **Filter and Sort** dialog box, choose the criteria you want to use for the filter. For example, to generate copies of your main document only for addresses that list Australia as the country/region, you would click Country or Region in the Field list, Equal to in the Comparison list, and Australia in the Compare to list.

The placeholders, such as address and greeting are called Mail Merge fields. Fields in Word correspond to the column headings in the data file that you select.

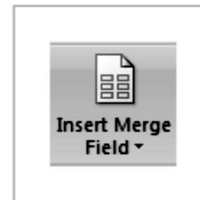
Columns in a data file represent categories of information. Fields that you add to the main document are placeholders for these categories. Rows in a data file represent records of information. Word 2007 generates a copy of the main document for each record when you perform a Mail Merge. By putting a field in your main document, you indicate that you want a certain category of information, such as name or address to appear in that location.

When you insert a Mail Merge field into the main document, the field name is always surrounded by chevrons (« »). These chevrons do not show up in the merged documents. They just help you distinguish the fields in the main document from the regular text. When you merge, information from the first row in the data file replaces the fields in your main document to create the first merged document. Information from the second row in the data file replaces the fields to create the second merged document, and so on.

7.10.3 Inserting Field Names into Master Document

You may wish to delete or add fields types to your Mail Merge document. You can do that easily. Just click on the **Customize Columns** button. The **Customize**

Columns dialog box opens. Then, click Add, Delete or Rename to alter the field types. You can also use the Move Up and Move Down buttons to rearrange the order of the fields. When you do, click on **OK**. Once you have added all your recipients, click on **OK** on the New Address List dialog box. Name the data source and click on Save. To insert a field into your document, click on **Insert Merge Field** in the Mailings option. Select the field you would like on to insert. The field name appears where you have the cursor located in your document. You can edit and format the text surrounding the field. Formats applied to the field will carry over to your finished document. You can continue to add insert fields to your document with the help of tool.



You may wish to delete or add fields types to your mail merge document. You can do that easily. Just click the **Customize Columns** button. The **Customize Columns** dialog box opens. Then, click on Add, Delete or Rename to alter the field types. You can also use the Move Up and Move Down buttons to rearrange the order of the fields. When you are done, click on **OK**. Once you have added all your recipients, click on **OK** on the **New Address List** dialog box. Name the data source and click on **Save**.

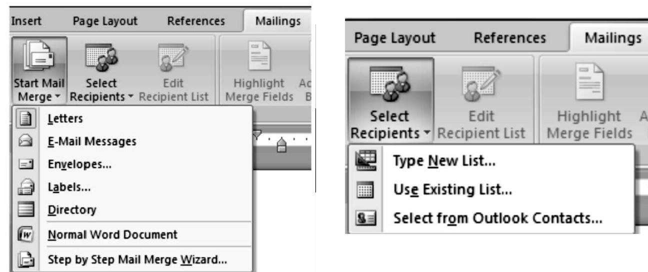


Using Address Book

You can use the **Insert Address** button to select an address from your contacts list. To add an address to your contacts list, you must add it to the address book in your e-mail program. On the **Tools** menu, point to **Letters and Mailings** and then click on **Envelopes and Labels**.

- Click on the **Envelopes** or **Labels** tab.
- Click on **Insert Address** .
- In the **Show Names from the** box, click on the desired address book or contact list.

- In the **Type Name or Select from list** box, enter a name or click a name in the list.



NOTES

Creating a Mail Merge Document

Click on **Start Mail Merge** on the **Mailings** options and select the type of document you would like to create. For example, you can select letters, envelopes or labels. Select **Step-by-Step Mail Merge Wizard** for more help creating your document.

Click on **Select Recipients** on the **Mailings** options to add recipients to the mailing. You can opt to create a new database of recipients. You can also opt to use an existing list or Outlook contacts.

In the **New Address List** Box, begin entering your contacts. You can use the Tab key to move between fields. Each set of fields is referred to as an entry. To add additional recipients, click on the **New Entry** button. To delete an entry, select it and click on **Delete Entry**. Click **on Yes** to confirm the deletion.

7.10.4 Merge the Information onto One File

Once you established the data source and inserted the merge fields in the main document, you are ready to perform the merge. You can merge to a new file which contains a customized version of the main document for each record in the data source or you can merge directly to a printer or e-mail message.

Sending the Merged Document to the Printer

The last step is to execute the merge operation. A form letter merge results in a separate letter for each record included from the data source. You can send the merge either directly to the printer or to a new document. If the appropriate software is installed, you also can send the merge to a fax or e-mail application. Select a destination from the dialog box accessed via the **Merge** button as shown in the screen.

7.10.5 Print the Merged Letters

One popular use of the Mail Merge feature is to create form letters. Mail Merge allows you to set up one letter, which serves as your main document. Using data from a table or external database, you can print this letter with different information for each record in the database or table. These instructions assume that you have

an understanding of the Word Mail Merge process. Following steps are required to print the merged letters:

Open a Blank Word Document

NOTES

- From the **Mailings** tab, in the **Start Mail Merge** group, click on **Start Mail Merge.** The **Start Mail Merge** sub-menu appears.
- From the **Start Mail Merge** sub-menu, select **Letters.**
- In the **Start Mail Merge** group, click on **Select Recipients and** select **Use Existing List...**

The **Select Data Source** dialog box appears.

- From the **Look in pull** down list, locate and select the file you will use for your list
- Click on **Open.**
- Following steps are required to select the recipients, you want to include in your Mail Merge.
  - o In the **Start Mail Merge** group, click on **Edit Recipient List.**
  - o Select the recipients.
  - o A recipient is selected if the checkbox beside their entry is selected.
  - o Click on **OK.**

To edit the recipient information, following steps are required:

- If you have not already done so, in your document, type the text of your letters.
- To insert merge fields,
  - o Position the insertion point at the proper place in your document.
  - o In the **Write and Insert Fields** group, click on **Insert Merge Field** and select the desired field.
- When finished, click on **Preview Results.**

A preview of your first recipient's letter appears. To preview your document with other recipients' information, in the **Preview Results** section, click on the arrows. To print the letters,

- Click **Finish & Merge** and select **Print Documents...**

The **Merge to Printer** dialog box appears.

- To print letters for all of your records, select **All.**
- To print a letter for only the record displayed, select **Current record.**
- To print letters for only certain records, type a range in the text boxes.
- Click on **OK.**
- The **Print** dialog box appears.
- You can make any necessary adjustment and Click on **OK**.

7.10.6 Secretarial Services and Mail Merge

Previously, secretarial services were related with skills similar to shorthand writing, typing and answering the telephone calls. Considering the new technology and innovations which have triggered the demands for a range of office services, the

secretarial work is no longer restricted to only the typing skills. At present, the responsibilities of secretaries have been expanded and the innovative technical gadgets, such as audio tape recorders, transcription equipments and computers have become the essential part of the secretary's office. Nowadays, simply answering of telephone calls is extended to replying and writing e-mails. Besides, the secretarial services are typically used for letter editing, mail merge, computer training, mailing list and database management, fax sending and receiving, and designing newsletters, brochures, flyers and promotional materials which are the unique desktop applications. The secretarial services have also incorporated desktop publishing, spreadsheet design, Web related services, phone-in dictation, database/ mailing list management, transcription, invoicing, bookkeeping and billing.

All the secretarial services are performed by a specialized person called secretary. Specifically, a secretary is a person employed to handle correspondence, keep files and to accomplish clerical work for another person or an organization. The sole responsibilities include to keeps records, take minutes of the meetings and answers correspondence. Hence, a secretary is an administrative assistant who supports management by means of various project management, communication and organizational skills. Besides, a secretary is an officer who deals with correspondence, admits new members, and organizes official meetings and events. A secretary has to perform numerous administrative duties. Traditionally, these duties were typically related to correspondence, such as the typing of letters, maintaining files of paper documents, etc. The advent of Word processing software has drastically reduced the time required by such duties of the secretary. Using Word processing software, the secretaries can efficiently handle and manage all the administrative particulars to even run a high level conference, arrange the typical meeting, take the minutes at meetings and prepare meeting documents for review. A high quality Word processing skill in software can be achieved preferably via a secretarial course. Besides, the secretary must have written communication, verbal communication, organizing, listening and co-operating skills.

Utilizing technology efficiently, such as a Word processor, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, etc., a secretary can keep a department running smoothly. A Word processor is accountable for typing letters, reports, forms, financial documents and entering information into a company database. A secretary's responsibility using Word processors is to draft, edit and revise documents to ensure that all documents are updated as per the company standard.

A secretary uses mail merge to create a set of documents that are fundamentally identical but where each document contains unique address elements. Using mail merge, a secretary can create a set of labels or envelopes where the return address is the same on all the labels or envelopes, but the destination address is unique on each one. Also, a secretary can create a set of form letters, e-mail messages or faxes. In mail merge, the basic content is the same in all the letters, messages or faxes, but each contains information that is specific to the individual recipient, such as name, address or some other piece of personal data.

A secretary can use the mail merge tool to generate address labels for the mailings list. The various features of mail merge and how to use mail merge is already discussed. You can work with both Microsoft Excel and Microsoft Word work together.

NOTES

How to Mail Merge Address Labels Using Excel and Word

First open the mail merge tool for generating address labels for the required mailing list. To do so, follow the steps given below:

1. Open Microsoft Word and from the main menu select **Tools** → **Letters and Mailings** → **Mail Merge**. This will open another window to the right of your screen.
2. Select the type of document on which you want to work. Choose '**Labels**' to prepare the address labels. Choose 'Next' at the bottom of the right hand screen.
3. Modify the layout of the document to '**Label Options**'. A screen will appear which allows selecting the size and style of labels. Specify as per your choice. Double click on the label that you want to work with.
4. Choose '**Select Recipients**' from a currently saved list or create a new one.
5. Select '**Browse**' and go to '**My Documents**'. Now search for the file that holds the address data. You can search for an Excel file also. Select an existing Excel document and click **OK**. Once you find the right file, click on '**Choose All**'.
6. Arrange all the labels by adding the proper merge codes. Choose from the options given at the right. If you select '**More Items**', then you will view the merge codes that matches the address file. Click on '**Update All Labels**' to add the words '**Next Record**' after the merge codes.
7. Click on '**Preview Your Labels**' and then on '**Complete the Merge**'. Select the option to '**Edit Individual Labels**' then select '**All**'. The complete set of address labels will be displayed on the screen.

7.10.7 Secretarial Letter Writing

A well-written official letter eliminates wordiness and helps the reader understand the message quickly and clearly. It creates goodwill, strengthens customer relationship and is a source of reliable, useful information. Hence, letters must be written effectively and conform to certain standards. A well-written letter works on the following principles:

- It is clear and concise.
- It uses words and expressions effectively.
- It uses correct spellings and grammar.
- It is neither too short nor very lengthy.
- It has a single idea concentrated in one letter.
- It is factually correct.
- It is courteous with a personal touch.
- It has relevant information.
- It uses the right e-mail ID/address.

All the related databases for a letter to be used as mail merge can be created and saved so that the relevant information can be sent to individual person/organization or group of persons/organizations.

NOTES

Layout of a Letter

While there is a tendency towards adopting an informal style, the established form of layout is the **fully blocked style** as shown. This avoids inconvenience and confusion, and saves time by not incorporating tabs and indents. A company's in-house style includes rules on the layout of its letters. For most letters, single line spacing is used as this gives the letter a compact look. In case the letter is very short, one may need to use double spacing.

The fully blocked style has the following characteristics:

(i) All typed entries including the date, inside name and address, subject line paragraph, complimentary close and signatures begin at the left-hand margin. They form a vertical line down the page.

(ii) There is complete absence of punctuation marks from the date, salutation, the complimentary close and the endline of the inside name and address.

Fully Indented Style

Name of the Company
Address & Phone No.
Date
Inside Name & Address
Salutation
Subject
Body of the Letter
Complimentary Close
Signature
Designation

A letter consists of the following essential parts:

(i) **Heading:** The heading shows the name and address of the organization, its telephone and fax numbers; the Internet and e-mail addresses, along with the logo of the company.

(ii) **Reference:** This could be a file number. It indicates the company's or the recipient's reference if it is a continuing correspondence. For example, Ref: Our letter No Pur/134/ 07 Re: Delay in receipt of goods.

(iii) **Subject Line:** This is often omitted, but its inclusion helps the reader quickly grasp what the letter is about. The subject heading is prefixed with the word Subject: or Sub: and is underlined to make it more prominent. For example, Subject: Request for quotation.

NOTES

(iv) **Date:** This is the date on which the letter was written. There are two ways of writing the date, the English style (17 June 2012) and the international format defined by ISO which uses a numerical date system in the format YYYY-MM-DD where,

- YYYY is the year [all the digits, i.e., 2012]
- MM is the month [01 (January) to 12 (December)]
- DD is the day [01 to 31]

For example, 17th June 2012 in this international format is written as 2012-06-17.

(v) **Name and Address of the Recipient:** This is mentioned in the letter. The name includes the title of the person (Mr, Ms, Dr).

(vi) **Opening Salutation:** This greets the recipient of the letter. The words generally depend upon the level of formality that the writer maintains with the addressee.

'Dear Sir' or 'Dear Madam' is a little old-fashioned and organizations now prefer to use a more personal approach, though 'Dear Sir' or 'Dear Madam' is still used in very formal correspondence.

(vii) **Body of the Letter:** This part contains the actual message/content of the letter. It consists of several paragraphs. Always leave a line space between paragraphs. The first paragraph is the introductory one, which may include a reference to a previous correspondence, a telephonic conversation, etc. The middle paragraphs constitute the main body and deal with the subject matter. The last paragraph is the concluding one and states the desired action.

(viii) **Complimentary Close:** It is a polite, formal way to end a letter. Standard forms are *Yours faithfully* or *Yours sincerely*. The complimentary close must match the salutation.

Dear Sir or Dear Madam matches with *Yours faithfully*

Dear Mr Smith matches with *Yours sincerely*

(ix) **Signature Space:** Leave space to add digital signature in the letter. The name and designation of the person signing the letter should be added below the signature for clarity.

(x) **Enclosures:** If some documents are being sent with the letter, they are indicated on the left-hand bottom corner, abbreviated as 'Encls' or 'Encl' (in case of a single document).

(xi) Lastly, the address of the 'registered office' and a registration number can be mentioned in the footer of the page. If required, the header can also be defined as per the need.

NOTES

Sample of Correspondence with Banks

Correspondence with banks is an indispensable part of correspondence for any organization. Banks finance new companies and are also a source of funds for expansion programs of existing companies. Transfer of money to suppliers, receipt of payment from customers or other financial transactions are done through the company's bankers. Most correspondence with banks is of routine nature, like the opening of an account, requesting a cheque book, availing an overdraft facility, requesting a stop payment, etc. Form letters for routine correspondence are available in most banks, though these are not always used. Letters serving other purposes also need to be written to banks. Here are some samples of correspondence with banks.

(i) Request for Opening of a Bank Account

The Manager

ICICI Bank

Lawrence Road

Amritsar

Dear Sir/ Madam

Subject: Request for Opening a New Current Account

We want to open a new current account in your bank in the name of our company 'XYZ Engineering Works Pvt. Ltd., Ghaziabad'. The following documents, as required for opening an account are forwarded, alongwith the application form enclosed with this letter:

- Certificate of Registrar of Companies granting permission to the company to commence business
- Letter of introduction by Mr B.R. Sharma, an account holder with your bank
- Certified copy of the proposal of the Board of Directors to open a Current Account duly signed by the Chairman

We are sending '15000/- in cash through Mr Amit Chopra as the initial amount to be deposited in the account. We would be grateful if you could kindly open the account in the name of the company.

Mr Anuj Adhikari, the Managing Director of our company, is authorized to operate this account. His specimen signatures are appended below.

You are requested to send us the cheque book, passbook and pay-in slip book to enable us to operate the account.

We are looking forward to a fruitful association with your bank.

With regards,

Yours sincerely,

Enclosures: Three

(ii) Request for account statement

NOTES

The Manager

ICICI Bank

Lawrence Road

Amritsar

Dear Sir/ Madam

Subject : Request for Current Account statement

We hold a Current Account in the name of our company 'XYZ Engineering Works Pvt Ltd., Ghaziabad' in your bank. Our account number is C 987345123. The account statement for the quarter January – March 2012 has not been received by us.

We would be grateful if you could send us the Current Account statement for the afore-mentioned period, i.e., 1 January – 31 March 2012, at the earliest to enable us to reconcile the bank balance.

With regards,

Yours sincerely,

Check Your Progress

1. What is the significance of Microsoft Office 2007?
2. What is a mini toolbar?
3. What is MS Word 2007? What can be done in MS Word 2007?
4. What is title bar?
5. How you can create a new document in MS Word 2007?
6. How the text can be inserted in a MS Word 2007 document?
7. Write the steps to add border and shading to MS Word 2007 document.
8. What is hyperlink? How a hyperlink is created?
9. How rows and columns are inserted in a table?
10. What is Watermark? How it is inserted into a MS Word 2007 document?
11. What is Mail Merge feature in MS Word 2007? Why it is used?

7.11 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Microsoft Office 2007 (codenamed Office 12) is a version of Microsoft Office, a family of office suites and productivity software for Windows, developed and published by Microsoft. It was preceded by Office 2003 and succeeded by Office 2010. Office 2007 introduced a new Graphical

NOTES

User Interface (GUI) called the Fluent User Interface (FUI), which uses ribbons and an Office menu instead of menu bars and toolbars. Office 2007 also introduced Office Open XML file formats as the default file formats in MS Excel, MS PowerPoint, and MS Word. The new formats are intended to facilitate the sharing of information between programs, improve security, reduce the size of documents, and enable new recovery scenarios.

2. The new Mini Toolbar is a small toolbar with basic formatting commands that appears within the document editing area, much like a context menu. When the mouse selects part of the text, Mini Toolbar appears close to selected text. It remains semi-transparent until the mouse pointer is hovered on it, to avoid obstructing what is underneath.
3. Microsoft Word 2007 is a product of Microsoft Office 2007 or 2007 Microsoft Office System. It is the latest version of Windows of the Microsoft Office System. It introduced the 'Ribbon User Interface' instead of the old menu and toolbar which is a work oriented Graphical User Interface (GUI).
4. The Title bar is next to the Quick Access toolbar. It displays the title of the current document which is in use. The first new document in Word is named as 'Document1' as shown below. When you open more new documents, Word automatically names them as Document2, Document3, Document4, etc., sequentially. The document can be saved by giving it a proper file name as per the user's choice.
5. To create a new document, Click on the **Microsoft Office Button** and then Click on **New** or Press **CTRL+N** on the keyboard.

   You will see that when you click on the Microsoft Office Button and then Click on **New**, Word provides number of choices about the types of documents you can create. Select and click on **Blank** from the list if you want to create a blank document. You can create a document using the option **Installed Templates**. Select any one of the templates as per your requirement. You can also browse other options through the list of choices that appear on the left.
6. Text can be inserted in a document at any point using any of the following methods:

   Type Text: Put your cursor where you want to add/insert the text and begin typing.

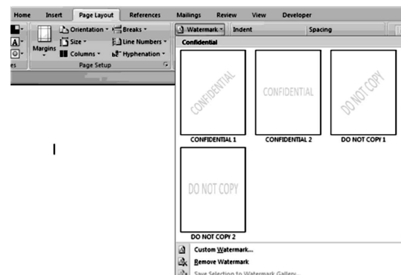
   Copy and Paste Text: Highlight the text you wish to copy and right click to view options. Now click on Copy option. Put your cursor where you want the text to be inserted in the document and right click to view options. Select Paste to paste the copied text.

   Cut and Paste Text: Highlight the text you wish to cut and right click to view options. Now click on Cut option. Put your cursor where you want the text to be inserted in the document and right click to view the options. Select Paste to paste the cut text.

NOTES

Drag Text: Highlight the text you wish to move. Click on it and drag it to the place where you want the text to be inserted in the document.

7. You can add borders, shade paragraphs and even entire pages. To create a border around a single paragraph or paragraphs, you need to do the following:
   - Select the area of text where you want to add the border or to add shade.
   - Click on the **borders** Button in the **Paragraph** Group on the **Home** Tab.
   - Select the **Border and Shading** option.
   - Select the options as per your choice and requirement.
8. A hyperlink or link is a reference to data that the reader can directly follow or that is followed automatically. A hyperlink points to a whole document or to a specific element within a document. Hypertext is text with hyperlinks. A user following hyperlinks is said to navigate or browse the hypertext. A hyperlink has an anchor, which is the location within a document from which the hyperlink can be followed.
9. To insert rows and columns, you need to select shortcut keys which provide options, such as Insert Columns to the Left, Insert Columns to the Right, Insert Rows Above and Insert Rows Below. After selecting the option, for example Insert Columns to the Right, you will get an extra column to the existing eight most column.
10. Watermark is a translucent image that appears behind the primary text in a document. To insert a watermark do the following:
    - Click on the **Page Lay**out Tab in the Ribbon.
    - Click on the **Watermark** button in the **Page Background** Group.
    - Click on the **Watermark** you want for the document or click on **Custom Watermark** and create your own watermark.
    - To remove a watermark, follow the steps above but click on **Remove Watermark.**



11. For example, in letters to your customers, each letter can be personalized to address each customer by name. The unique information in each letter comes from entries in a data source.

7.12 SUMMARY

NOTES

- Microsoft Office 2007 (codenamed Office 12) is a version of Microsoft Office, a family of office suites and productivity software for Windows, developed and published by Microsoft. It was preceded by Office 2003 and succeeded by Office 2010.
- Office 2007 introduced a new Graphical User Interface (GUI) called the Fluent User Interface (FUI), which uses ribbons and an Office menu instead of menu bars and toolbars. Office 2007 also introduced Office Open XML file formats as the default file formats in MS Excel, MS PowerPoint, and MS Word. The new formats are intended to facilitate the sharing of information between programs, improve security, reduce the size of documents, and enable new recovery scenarios.
- The Office 2007 button, located on the top-left of the window, replaces the File menu and provides access to functionality common across all Office applications, including opening, saving, printing, and sharing a file. It can also close the application. Users can also choose colour schemes for the interface.
- The ribbon, a panel that houses a fixed arrangement of command buttons and icons, organizes commands as a set of tabs, each grouping relevant commands. The ribbon is present in Microsoft Word 2007, Microsoft Excel 2007, Microsoft PowerPoint 2007, Microsoft Access 2007 and some Microsoft Outlook 2007 windows.
- Microsoft Word 2007 is a product of Microsoft Office 2007 or 2007 Microsoft Office System. It is the latest version of Windows of the Microsoft Office System. It introduced the 'Ribbon User Interface' instead of the old menu and toolbar which is a work oriented Graphical User Interface (GUI).
- The Microsoft Office button is located in the upper left corner of the Word 2007 Window. A menu appears when you click on this button. This menu helps in creating a new document or file, opening an existing document or file, saving a document or file, printing a document or file, sending the document or file via fax or e-mail, etc.
- The Quick Access toolbar is right next to the Microsoft Office button. This toolbar helps you to access the frequently used commands. The default commands which appear on this toolbar are Save, Undo and Redo. These commands help you to Save a document or file, Undo an action and Redo an action.
- The Ribbon is positioned at the top of the screen of the Word window. It includes eight tabs, namely Home, Insert, Page Layout, References, Mailings, Review, View and Add-Ins as shown in screen below. Each tab contains various new and advanced features of Word.
- Draft view is the most frequently used view. The Draft view is used for the purpose of editing a document quickly.

NOTES

- The Web Layout view enables you to see your document as it would appear in a browser such as Internet Explorer.
- The Print Layout view shows how the document will look when it is printed.
- The Reading Layout view formats your screen so that you can can read the document more comfortably. You can choose more such reading options that appear on the right side of the user interface.
- The Outline view displays the document in an outline format. You can display headings without the text. If you move a heading, the accompanying text also moves with it.
- Multiple documents can be simultaneously opened when you need to type or edit multiple documents at once. All the documents as opened will be listed in the **View Tab** of the Ribbon when you will click on Switch Windows.
- To enter text, just type the text in the Word window. The text will appear at the location of the blinking cursor. You can move the cursor using the arrow keys on the keyboard or by positioning the mouse and clicking the left button.
- The find and replace option can be accessed by selecting and pressing key combinations **CTRL+F** or **CTRL+H**.
- To print the document you need to check the options on the Ribbon. It tell you that now you are in the 'Preview' mode. You can change margin size, page orientation, print options and can select many other print functions in the 'Preview' mode.
- Print Preview option is used to see how the document will look like when it is printed on paper. You can get into Print Preview mode by clicking the **Print Preview** icon on the Quick Access Toolbar or clicking on the **Office Start Menu → Print → Print Preview**.
- Document formatting represents all the types of formatting part which are applied to a selected document for making suitable presentation of the Word Document.
- Font styles are predefined formatting options which are used to emphasize specific text. They include bold, italics and underline.
- Paragraph alignment helps in organizing the text according to the way you want it to appear.
- Bulleted lists are represented using bullet points, numbered lists are represented using numbers and outline lists combine together the numbers and letters depending on the association of the list.
- You can insert contents from another file, such as linking to a location in a document or Web page that you created in Word 2007. You must mark the hyperlink location or destination and then add the link to it. A hyperlink is a graphic or a piece of text that links to another Web page.

NOTES

- Tables organize data into rows and columns. They are used for the purpose of displaying data in a tabular format.
- Borders and shading facilitate you to highlight information in a table. Thus it enhances the appearance of the text in a table. Word automatically applies a ½ point border and a white background, i.e., no shading to all tables and table cells.
- In Word 2007, you can modify objects and include graphics in many ways to add sparkle to even the most mundane document. You can add drop caps, group two or more objects into one object and change the layering of overlapping objects.
- Mail Merge is one of the effective features of Microsoft Word. It helps you to draft and send letters to many recipients in very little time. You can use mail merge when you want to create a set of documents, such as a form letter that is sent to many customers. Each letter has the same kind of information yet the content is unique.
- To merge information into your main document, you must connect the document to a data source or a data file. If you do not already have a data file, you can create one during the Mail Merge process. If you use an existing list, make sure that it contains the information that you want to use, including all the columns and the rows.
- The placeholders, such as address and greeting are called Mail Merge fields. Fields in Word correspond to the column headings in the data file that you select.

7.13 KEY WORDS

- **User Interface (UI):** The new User Interface (UI), officially known as Fluent User Interface (FUI), has been implemented in the core Microsoft Office applications: Word, Excel, PowerPoint, Access, and Outlook. The default font used in this edition is Calibri.
- **Office button:** The Office 2007 button, located on the top-left of the window, replaces the File menu and provides access to functionality common across all Office applications, including opening, saving, printing, and sharing a file. It can also close the application. Users can also choose colour schemes for the interface.
- **Ribbon:** The ribbon, a panel that houses a fixed arrangement of command buttons and icons, organizes commands as a set of tabs, each grouping relevant commands.
- **Contextual tabs:** Some tabs, called Contextual Tabs, appear only when certain objects are selected. Contextual Tabs expose functionality specific only to the object with focus.

NOTES

- **Title bar:** A toolbar used to display the title of the current document which is in use.
- **Draft view:** A view used for the purpose of editing a document quickly.
- **Special characters:** Punctuation, spacing or typographical characters which are generally not available on the standard keyboard.
- **Paragraph alignment:** This option helps in organizing the text according to the way you want it to appear.
- **Justify:** An option which aligns the text to both the left and right margins.
- **Data source:** A file that contains the information to be merged into a document.

7.14 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What are new features of Microsoft Office 2007?
2. What is word processing?
3. What does Microsoft Office button menu provides?
4. How a blank page is inserted in a document?
5. Which button is selected if you want to check the margin, paper size and orientation to print the Word document?
6. How is graphic inserted in a MS Word document?
7. How is a nested list created?
8. How new styles are created?
9. Why 'Print' option is selected in mail merge operation?

Long-Answer Questions

1. Briefly discuss the significant features of Microsoft Office 2007.
2. Explain briefly the concept of 'getting started' and also name the various ways in which a document can be viewed in MS Word 2007.
3. Explain the process to create and edit document with the help of examples.
4. Explain the steps required in formatting a document.
5. Explain the steps required in creating a custom dictionary with the help of examples.
6. Discuss the features of spelling and grammar tool with the help of examples.
7. Discuss the steps required to use a Print function in Microsoft Word 2007.

8. Write all the steps required to insert symbols and graphics in MS Word document.
9. Describe the styles with reference to advanced document formatting.
10. Explain all the operations that can be performed in MS Word 2007 tables.
11. Explain the significance and working of Mail Merge with the help of examples.

7.15 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 8 MS EXCEL SKILLS

NOTES

Structure

8.0 INTRODUCTION

Microsoft Excel 2007 is an electronic spreadsheet that runs on a personal computer. A workbook is the Microsoft Excel 2007 file in which you enter and store related data. A worksheet, also known as a spreadsheet, is a collection of cells on a single sheet where you can actually keep and manipulate the data. You can use it to organize the data into rows and columns. You will learn use it to perform mathematical calculations quickly. You will also learn to create a spreadsheet using various formatting and editing features given in the Ribbon. The Ribbon is the panel at the top portion of the spreadsheet. You will also learn the concept of selection in a worksheet in which various operations, such as cell selection, row and column selection, entire worksheet selection and data range selection are performed to complete the task. Working with formula is the prime feature of Microsoft MS Excel 2007. The various mathematical formulas, such as SUM(), PI(), POWER(), FACT(); date functions, such as NOW(), DATE(), TODAY(),

NOTES

DAY(), YEAR(); logical functions, such as IF(), AND(), OR(), NOT(); text functions, such as LOWER(), PROPER(), LEN(), LEFT(), MID() and TRIM() help to work with MS Excel 2007. You can create graphs and charts in the worksheet by inserting specified Charts, Clip Art, SmartArt, Shapes and pictures to enhance your worksheet. MS Excel 2007 supports charts, graphs or histograms generated from specified groups of cells. The generated graphic component can either be embedded within the current sheet or added as a separate object. The generated graphic component can either be embedded within the current sheet or added as a separate object. OLE or Object Linking and Embedding allow a Windows application to format or calculate data. You can freeze and unfreeze rows and columns and can also hide or unhide any worksheet. You will learn the various steps to sort and filter data, and generate the result as per your requirement.

In this unit, you will study about the basics of MS Excel skills, entering and editing cell entries, working with numbers, changing the worksheet layout, formatting text, borders and colour, printing in MS Excel, using functions and references in MS Excel, and naming ranges in MS Excel.

8.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the significance of Microsoft Excel 2007
- Explain selection process in a worksheet
- Work with formulae
- Edit a worksheet
- Create charts in MS Excel 2007
- Create and execute macro
- Print worksheets data in MS Excel 2007

8.2 MICROSOFT EXCEL 2007

Microsoft Excel 2007 is the newest version of spreadsheet software in the Microsoft 2007 Office Suite. This spreadsheet application is specifically developed by Microsoft for Microsoft Windows and Mac OS X. Technically speaking, MS Excel 2007 files are referred as spreadsheets. This is a generic term, which sometimes means a workbook (file) and sometimes means a worksheet (a page within the file). Data files created with MS Excel 2007 are called workbooks. MS Excel 2007 files by default contain three blank worksheets. This gives you the flexibility to store related data in different locations within the same file. More worksheets can be added and unwanted worksheets can be deleted as per the user requirement. Thus, MS Excel 2007 is a powerful and most extensively used tool as spreadsheet application which allows you to store, organize and analyse numerical, graphic and text data. Spreadsheets allow information to be organized in rows and tables, and can be analysed using various mathematical, trigonometric, text, logical, date and time functions. The number of rows is now 1,048,576 (2^{20})

NOTES

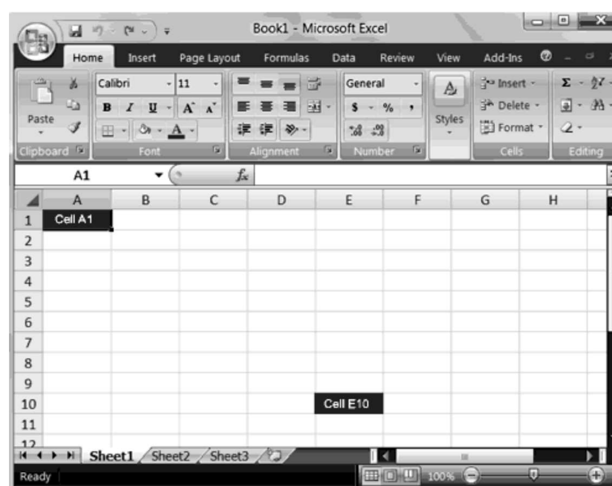
and columns is 16,384 (2^{14}). Microsoft Excel 2007 has the basic features of all spreadsheets, using a grid of cells arranged in numbered rows and letter named columns to organize data manipulations. Further with Microsoft Excel 2007 you can analyse, manage and share information quickly and easily to formulate more knowledgeable decisions. With the new user interface, rich data visualization and PivotTable views, professional looking charts can be created easily.

The advanced features in Microsoft Excel 2007 include Office themes, more styles, rich conditional formatting, easy formula writing, Sort & Filter, data validation, worksheet and workbook protection, Goal Seek, Scenario, PivotTable and PivotChart. Goal Seek and Scenario are part of What-If Analysis tools. MS Excel 2007 supports charts, graphs or histograms generated from specified groups of cells. The generated graphic component can either be embedded within the current sheet or added as a separate object. OLE or Object Linking and Embedding allow a Windows application to format or calculate data. This may acquire the form of 'embedding' where an application uses another to handle the task, for example a MS PowerPoint presentation can be embedded in an MS Excel 2007 spreadsheet or vice versa.

You can create a spreadsheet using various formatting and editing features given in the Ribbon panel. Thus, you can perform calculations using the functions given in MS Excel 2007 and can sort and filter data as per your requirement. You can create graphs in the worksheet, insert illustrations, Clip Art, SmartArt, Shapes and pictures to enhance your worksheet. You can freeze and unfreeze rows and columns and can also hide and unhide any worksheet.

8.2.1 Worksheet, Workbook and Workspace

A Microsoft Excel 2007 file in which you can enter and store related data is known as a workbook. A workbook is also identified as a spreadsheet that is a group of cells on a single sheet where you in fact keep and operate data. Every worksheet consists of columns and rows. The columns are lettered A to Z and then continue with AA, AB, AC, and so on. The rows are numbered from1 to 1,048,576. The number of rows and columns that you can hold in your worksheet is restricted by computer memory and your system resource.



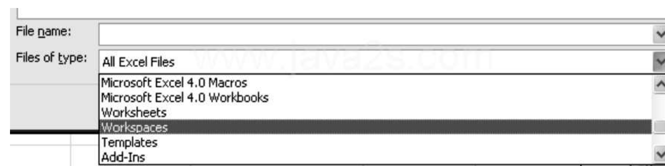
NOTES

Cell address is the combination of a column letter and a row number. For example, if a cell is located in the upper left corner of the worksheet, which is A1, this means that it is located in column A and row 1. Similarly, cell E10 is situated in column E on row 10. The data can be entered into the cells present on the worksheet. N-number of worksheets can be present in a workbook. To work with workspace you have to perform the following steps:

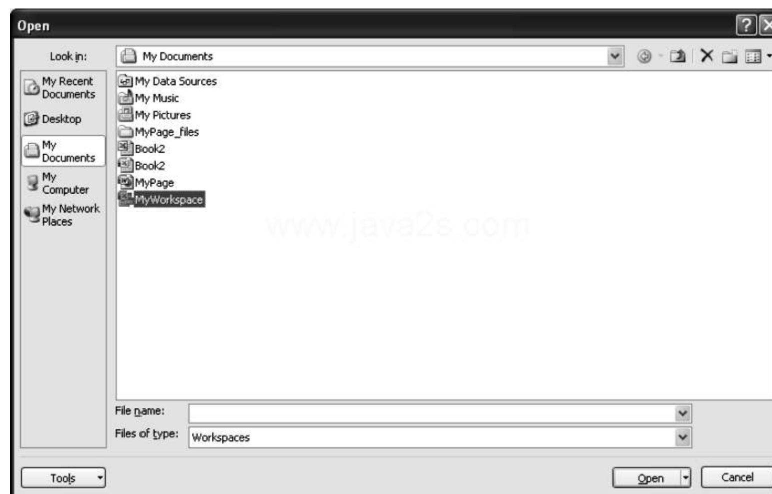
- Click the Office button. Click **Open**.



- Click the **Files of type:** list arrow. Click **Workspace**.



- Select the workspace file. Click **Open**.



8.2.2 Getting Started

As soon as you start MS Excel 2007, you will see that its features are similar to the previous versions. You will also see that there are many additional features which help you to work with special effects. Three new features that are included in MS Excel 2007 are the Microsoft Office Button, the Quick Access Toolbar and the Ribbon.

NOTES

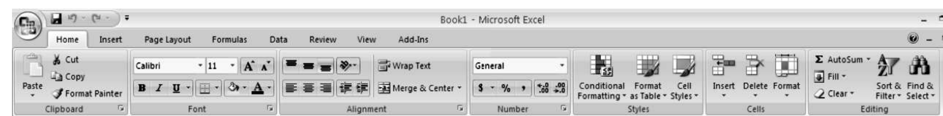
Microsoft Office Button

The **Microsoft Office Button** performs various functions that were found in the File menu of older versions of MS Excel 2007. This button permits you to create a new workbook, open an existing workbook, save a workbook using Save and Save As, Print, Send or Close a workbook.

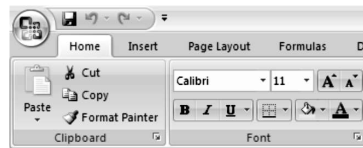


Ribbon

The **Ribbon** is the panel at the top portion of the spreadsheet. It includes seven tabs namely, **Home, Insert, Page Layout, Formulas, Data, Review** and **View**. **Add-ins** is another option which is automatically displayed on the Ribbon when you add any new application to the program. Each tab is a collection of features designed to perform specific functions that you require while creating or editing MS Excel 2007 spreadsheets.



The frequently used features are displayed on the Ribbon. To view additional features of each group, click on the arrow at the bottom right corner of each group.

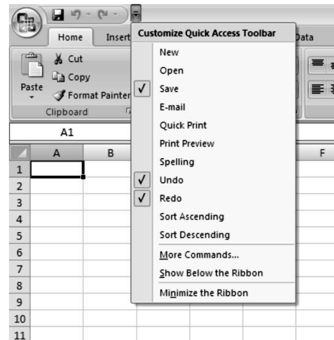


- **Home:** Clipboard, Fonts, Alignment, Number, Styles, Cells, Editing.
- **Insert:** Tables, Illustrations, Charts, Links, Text.
- **Page Layouts:** Themes, Page Setup, Scale to Fit, Sheet Options, Arrange.
- **Formulas:** Function Library, Defined Names, Formula Auditing, Calculation.
- **Data:** Get External Data, Connections, Sort & Filter, Data Tools, Outline.
- **Review:** Proofing, Comments, Changes.
- **View:** Workbook Views, Show/Hide, Zoom, Window, Macros.

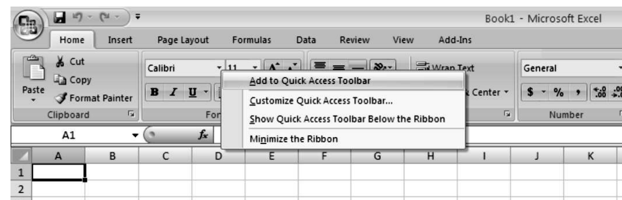
Quick Access Toolbar

The **Quick Access Toolbar** can be customized as per the user need and contains commands that you use most frequently. You can place the Quick Access Toolbar

above or below the Ribbon. To change the location of the Quick Access Toolbar, click on the arrow at the end of the toolbar and click **Show Below** the Ribbon.



You can also add more items to the Quick Access toolbar. To do this, right click on any item in the Office Button or the Ribbon and then click on Add to Quick Access Toolbar. A shortcut will be added there.



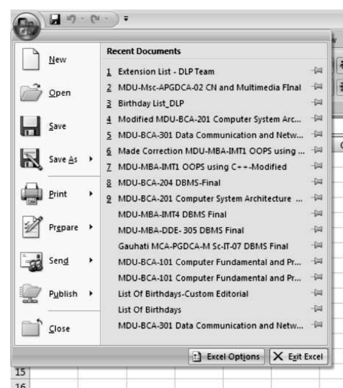
Mini Toolbar

Mini Toolbar is a new feature in Microsoft Office 2007. This is a floating toolbar and is displayed when you select text or right click any text. It displays the common formatting tools, such as Bold, Italic, Fonts, Font Size and Font Color.



MS Excel 2007 Options

MS Excel 2007 provides a wide range of customizable options that help you to create an MS Excel 2007 workbook of required specifications. To access these customizable options, follow the given steps:



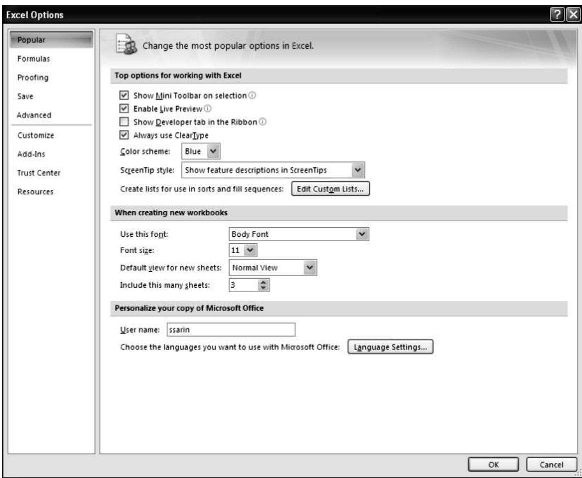
NOTES

- Click on the Office Button.
- Click on MS Excel 2007 Options which you will get from Quick Access Toolbar.

NOTES

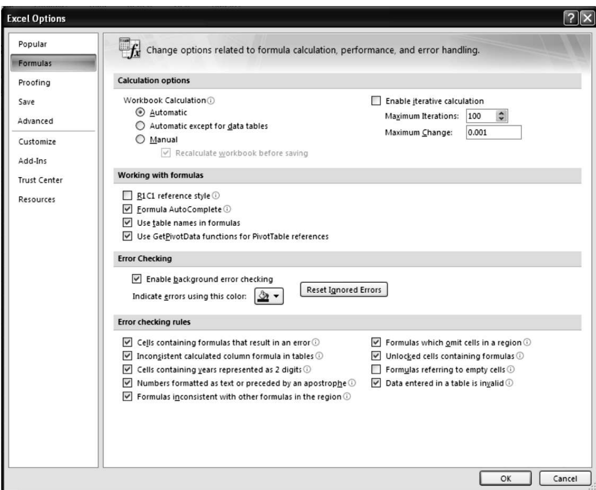
Popular

The **Popular** features helps you to personalize your work environment using the Mini Toolbar, Color schemes, default options when creating new workbooks and creating lists for sort and fill sequences. It also helps you to access the Live Preview feature to preview how a feature affects the document as you hover over different choices. The choices provide new font size, table style or cell style which can be applied on a workbook as per requirement.



Formulas

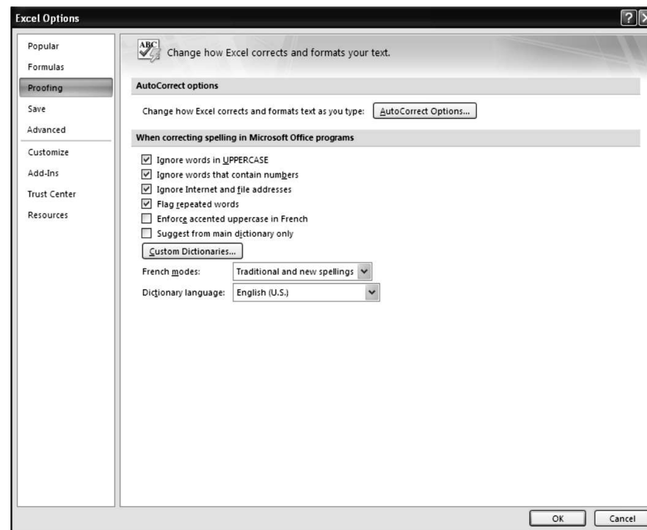
The **Formulas** feature permits you to modify the calculation options, to work with formulas, error checking and error checking rules. Working with formulas provide four check boxes which are R1C1 reference style, Formula AutoComplete, Use table names in formulas and Use GetPivotData functions for PivotTable references as shown in the given screen.



NOTES

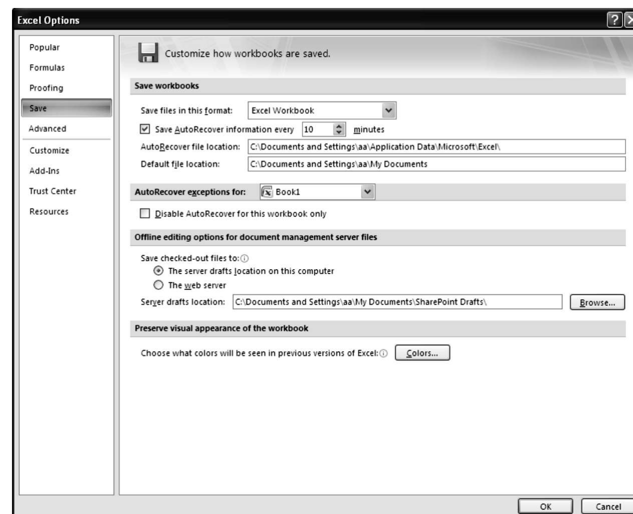
Proofing

The **Proofing** feature permits you to personalize the options for correcting words and formats of your text. You will get AutoCorrect option in Proofing feature. You can customize auto correction settings so that it will ignore certain words or errors in a document via the **Custom Dictionaries**...



Save

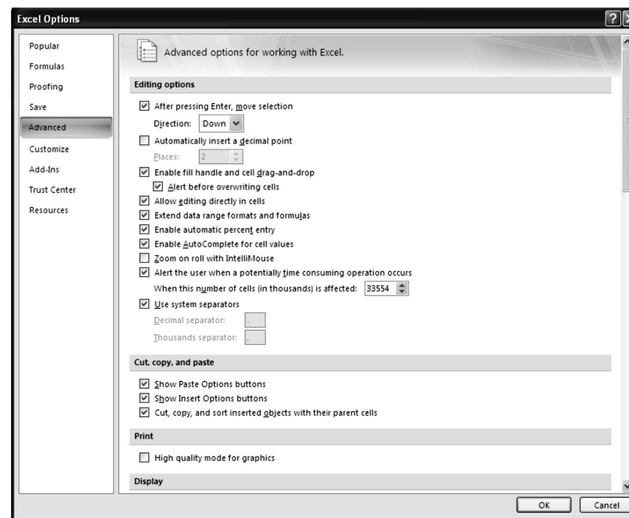
The **Save** feature permits you to personalize your workbook when saved. You can also specify how often you want auto save to run and where to save the workbooks.



Advanced

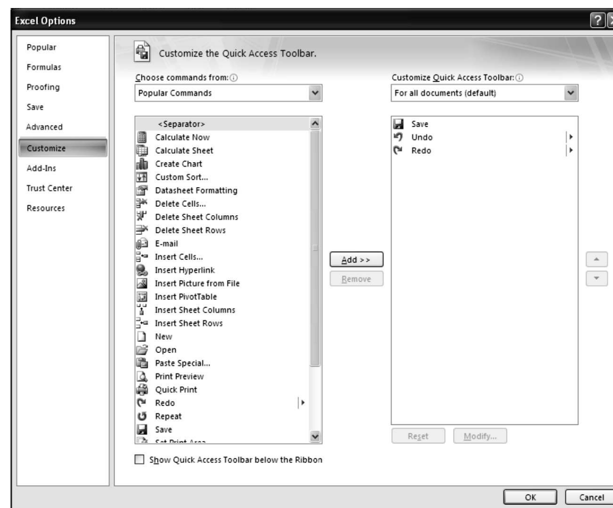
The **Advanced** feature permits you to specify the options for editing, copying, pasting, printing as well as displaying formulas, calculations and other general settings.

NOTES



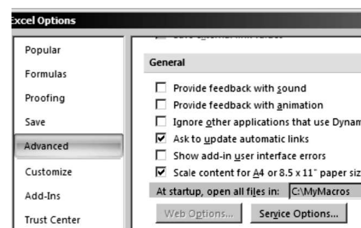
Customize

Customize permits you to add specific features to the Quick Access Toolbar. It adds the tools which you frequently use.



Opening MS Excel 2007 Application

To open MS Excel 2007 application, following steps are required:



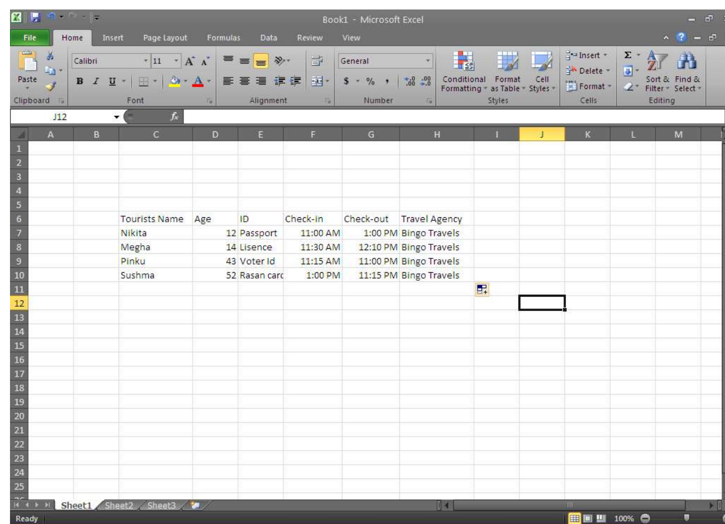
- Click the Office Button, then press right to select **Customize Quick Access Toolbar** which provides MS Excel 2007 Options tab.

NOTES

- Click the Advanced category and scroll down up to the General section.
- In the box for '**At startup, open all files in**', you might see the name of a folder and its path. Clear the folder information from that box or go to that folder and remove the unwanted files. Click **OK** to close the MS Excel 2007 Options dialog box.

Entering Information in a Worksheet

To enter information in a worksheet, you need to open an empty workbook and enter the data as shown in the screen, below:



Moving Around Worksheet and Workbook

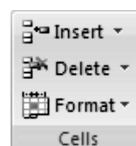
The arrow keys give you the option to move around your worksheet. With the help of down arrow key you can move downward one cell at a time. Similarly, the up arrow key can be used to move upward one cell at a time. You can even move across the page to the right, one cell at a time by using the Tab key. By holding hold the SHIFT key and then pressing the Tab key, you can move to the left, one cell at a time. You even have the right and left arrow keys available by which you can also move right and left respectively, one cell at a time. The Page Up (Pg Up) and Page Down (Pg Dn) keys move up and down one page at a time. By pressing down the CTRL key and simultaneously pressing the Home key, you can move to the beginning of the worksheet.

It is convenient to either move or copy the whole worksheet. The term worksheet refers to the main document that you use in MS Excel 2007 to store and work with data. There might be a chance that calculations or charts that are based on worksheet data might turn out to be inaccurate if you shift the worksheet. You can move or copy worksheet by inserting between sheets that are referred by a 3-D reference. This reference refers to a range that spans two or more worksheets in a workbook. Data on that worksheet might be unexpectedly included in the calculation. Select the worksheets that you want to move or copy as shown in the screen below.



NOTES

To move to the next or previous sheet tab, you can also press CTRL + Pg Up or CTRL + Pg Dn. On the **Home** tab, in the **Cells** group, click **Format** and then under **Organize Sheets**, click **Move or Copy Sheet**.



You can also right click a selected sheet tab and then click **Move or Copy**. In the **Move or Copy** dialog box, in the **Before sheet** list, do one of the following:

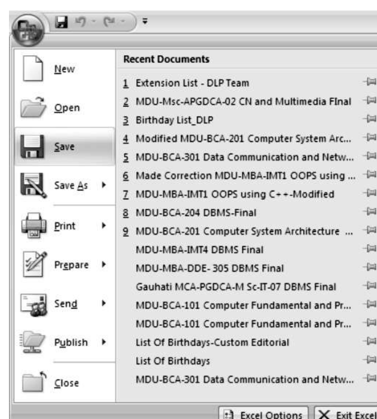
- Click the sheet before which you want to insert the moved or copied sheets.
- Click **move to end** to insert the moved or copied sheets after the last sheet in the workbook and before the **Insert Worksheet** tab.



To copy the sheets instead of moving them, in the **Move or Copy** dialog box, select the **Create a copy** check box.

Saving a Workbook

To save a workbook, you have two options, **Save** and **Save As**. To save a document, follow the given steps:

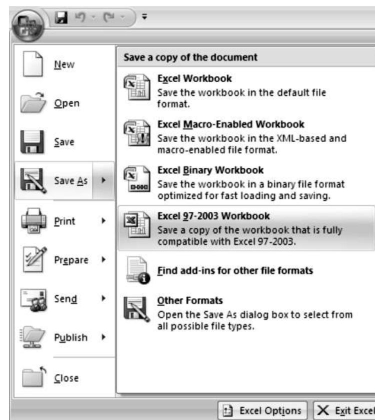


- Click on the **Microsoft Office Button.**
- Click on **Save**.

You can also use the **Save As** feature to save the workbook with a different name or to save it as earlier versions of MS Excel 2007. The older versions of MS Excel 2007 cannot be opened in an MS Excel 2007 worksheet unless you save it as an MS Excel 97-2003 Format. To use the **Save As** feature, follow the given steps:

- Click on the **Microsoft Office** button.
- Click on **Save As.**

- Give a name for the workbook.
- In the **Save as Type** box, select **Excel 97-2003** workbook.

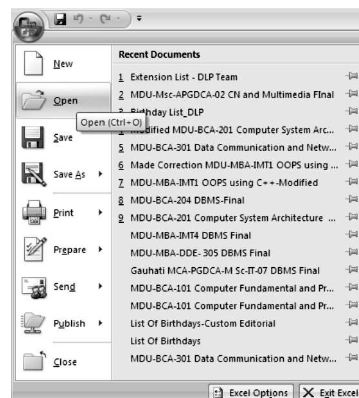


Closing a Workbook File

To close a workbook file, you need to press CTRL+F4 key combination. You can close all open workbooks wihout closing MS Excel 2007. For this, you need to open file menu and select Close All option.

Opening an Existing Workbook File

To open an existing workbook, follow the given steps:



- Click on the **Microsoft Office** button.
- Click on **Open.**
- Browse to the workbook.
- Click on the title of the workbook.
- Click on **Open**.

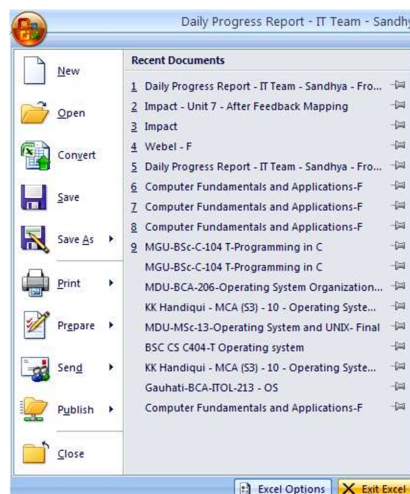
NOTES

Quitting From MS Excel 2007

To quit from MS Excel 2007, click on **Microsoft Office Button** and then select Exit MS Excel 2007 button.

NOTES

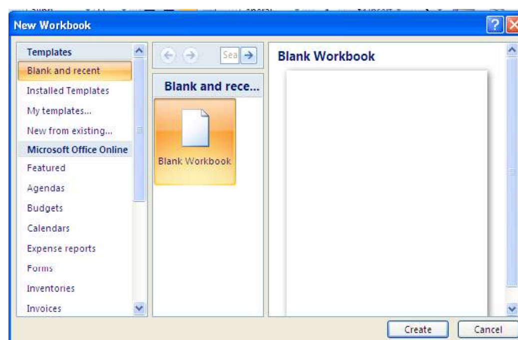
You will quit from MS Excel 2007.



8.3 CREATING AND EDITING WORKSHEET

You can create a new worksheet and can edit any existing worksheet as per your requirements. To create a worksheet, follow the given steps:

- Click on the **Office Button** and select **New** option. It will open the **New Workbook** dialog box as shown below. From the **Templates** group select **'Blank and recent'** option.
- Now click on **'Blank Workbook'** and select Create button. A new blank spreadsheet will be displayed on the screen. Give a name to this worksheet and save the file with '.xlsx' file name extension.

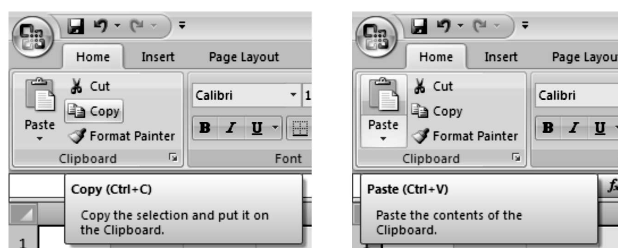


You can edit a worksheet if any incorrect information has been entered into a cell. To do so, click on the cell and enter the correct information. Typing replaces whatever is in the cell. Editing a spreadsheet includes copy, cut, paste, move, changing column width/row height, cell alignment and formatting, font and number formatting, inserting and deleting cell(s)/row(s)/column(s), insert/copy/move/rename/delete a worksheet.

Copy and Paste

To copy and paste data, follow the given steps:

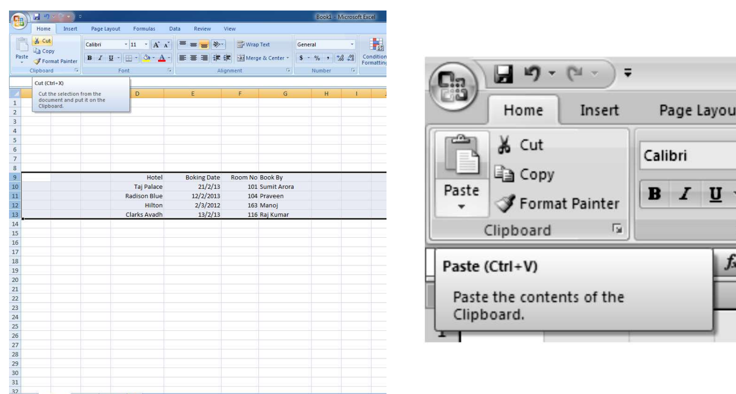
- Select the cell(s) that you want to copy.
- On the **Clipboard** group of the **Home** tab, click on **Copy.**
- Select the cell(s) where you want to copy the data.
- On the **Clipboard** group of the **Home** tab, click on **Paste.**



Cut and Paste

To cut and paste data, follow the given steps:

- Select the cell(s) that you want to copy.
- On the **Clipboard** group of the **Home** tab, click on **Cut.**
- Select the cell(s) where you want to copy the data.
- On the **Clipboard** group of the **Home** tab, click on **Paste.**



Undo and Redo

To undo or redo the most recent actions, click on **Undo** or **Redo** on the **Quick Access Toolbar**.



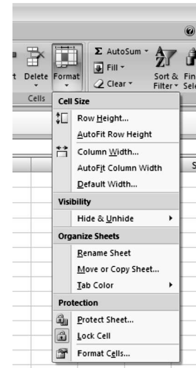
8.3.1 Changing Column Width/Row Height

In MS Excel 2007, you have the option to modify any column width or row height in your worksheets so that the readability and look of the data can be changed. For example, if your worksheet contains large numbers, you can enlarge the columns to make the worksheet less cluttered. It is much more advisable to expand the

NOTES

NOTES

width of the columns that contain cells with truncated text entries or numbers that MS Excel 2007 shows as ######.



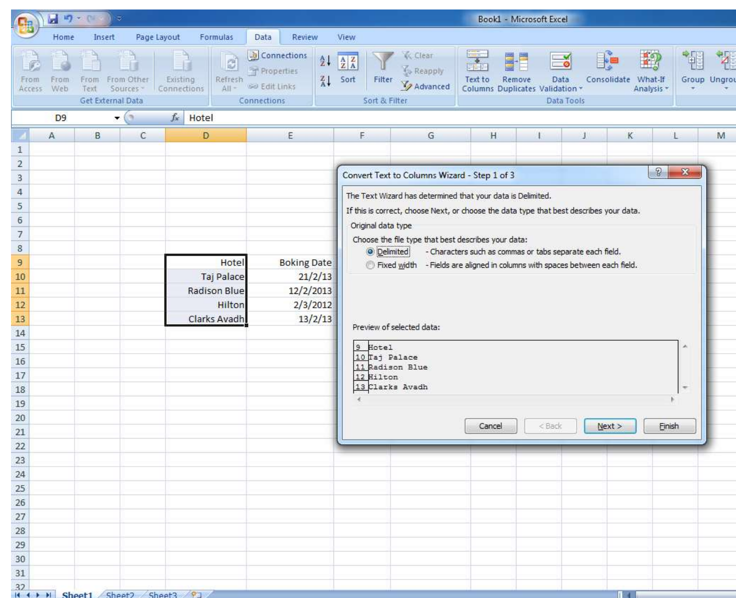
To change the width of a column or the height of a row, follow the given steps:

- Click on the **Format** button on the **Cells** group of the **Home** tab.
- Manually adjust the height and width by clicking on **Row Height** or **Column Width.**
- To use **AutoFit,** click on **AutoFit Row Height** or **AutoFit Column Width**.

8.3.2 Cell Formatting—Font, Alignment and Number

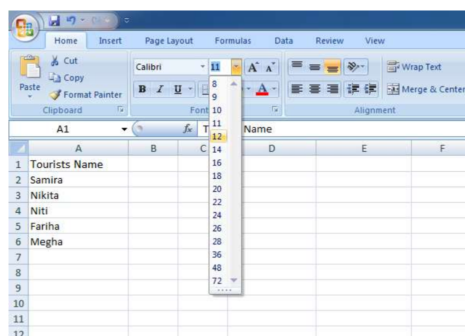
Convert Text to Columns: Sometimes, you may have to split data of one cell and place them in two or more cells. You can do this by utilizing the Convert Text to Columns Wizard.

- Highlight the column in which you wish to split the data.
- Click on the **Text to Columns** button on the **Data** tab.
- Click on **Delimited radio button** if you have a comma or tab separating the data or click fixed widths to set the data separation at a specific size.



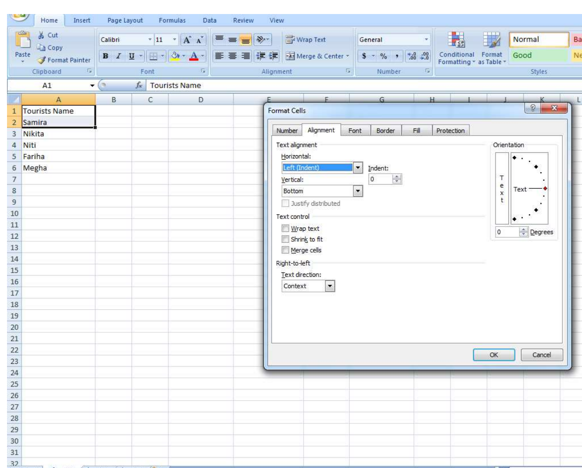
NOTES

Modify Fonts: Modifying fonts in MS Excel 2007 helps you to emphasize titles and headings. To modify a font, follow the given steps:



- Select the cell or cells where you want to apply the font.
- On the **Font** group on the **Home** tab, select the font type, size, bold, italics, underline or color.

Format Cells Dialog Box: In MS Excel 2007, you can apply specific formatting to any cell. To apply formatting to a cell or group of cells, follow the given steps:



- Select the cell or cells to change the formatting.
- Click on the **Dialog Box** arrow on the **Alignment** group of the **Home** tab.

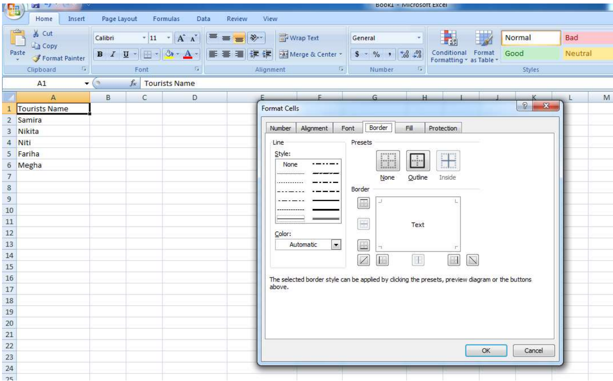
There are various tabs on this dialog box that permit you to modify the properties of the cell or cells.

- **Number:** Allows the display of different number types and decimal places.
- **Alignment:** Allows the horizontal and vertical alignment of text, wrap text, shrink text, merge cells and the direction of the text.
- **Font**: Allows control of font, font style, size, color and additional features.
- **Border:** Changes border styles and colors.
- **Fill:** Fill colors and styles.
- **Protection:** Locking cells and hiding formulas.

NOTES

Adding Borders and Colors to Cells

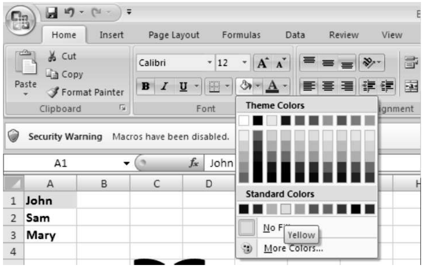
Borders and colors can be added to cells manually or using the styles. To add borders manually, follow the given steps:



- Click on the **Borders** drop-down menu on the **Font** group of the **Home** tab.
- Select the suitable border.

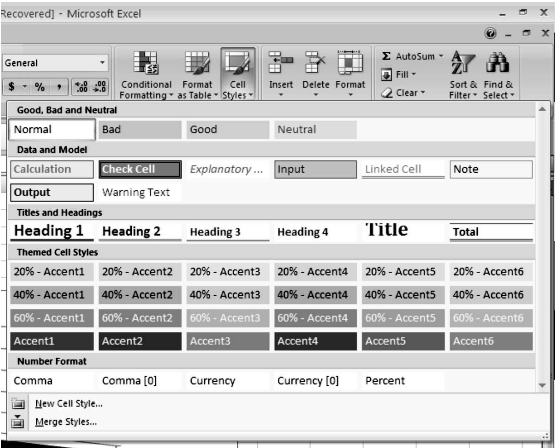
To apply colors manually, follow the given steps:

- Click on the **Fill** drop-down menu on the **Font** group of the **Home** tab.
- Select the proper color.



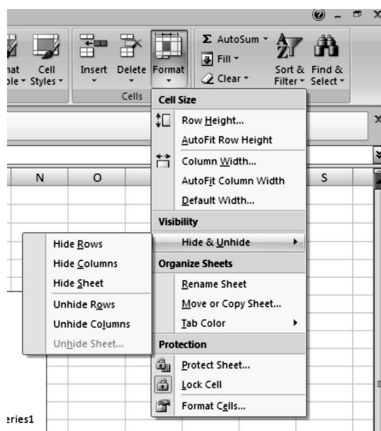
For applying borders and colors using styles, follow the given steps:

- Click on **Cell** Styles on the **Home** tab.
- Select a style or click on **New Cell Style**.



Hide or Unhide Rows or Columns: To hide or unhide rows or columns, follow the given steps:

- Select the row or column you want to hide or unhide.
- Click on the **Format** button on the **Cells** group of the **Home** tab.
- Click on **Hide & Unhide**.



NOTES

8.3.3 Inserting and Deleting Cell(s)/Row(s), Column(s)

Inserting Cells, Rows and Columns

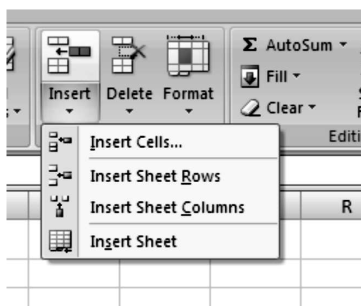
To insert cells, rows and columns in MS Excel 2007, follow the given steps:

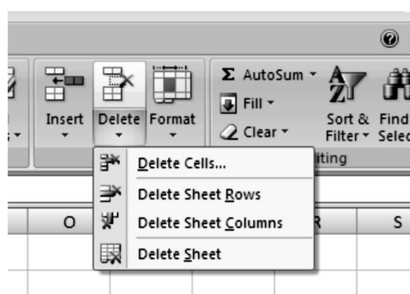
- Place the cursor in the row below where you want the new row or in the column to the left of where you want to have the new column.
- Click on the **Insert** button on the **Cells** group of the **Home** tab.
- Click on the proper option **Cell, Row** or **Column.**

Delete Cells, Rows and Columns

To delete cells, rows and columns, follow the given steps:

- Place the cursor in the cell, row or column that you want to delete.
- Click on the **Delete** button on the **Cells** group of the **Home** tab.
- Click on the appropriate option **Cell, Row** or **Column.**





NOTES

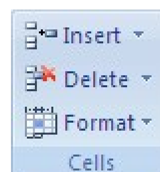
8.3.4 Insert/Copy/Move/Rename/Delete Worksheet

To insert a new worksheet, do any one of the following steps:

- To quickly insert a new worksheet at the end of the existing worksheets, click the **Insert Worksheet** tab at the bottom of the screen.



- To insert a new worksheet in front of an existing worksheet, select that worksheet and then, on the **Home** tab, in the **Cells** group, click **Insert** and then click **Insert Sheet**.



You can also right click the tab of an existing worksheet and then click **Insert**. On the **General** tab, click **Worksheet**, and then click **OK**.

Insert Multiple Worksheets at the Same Time

Hold down SHIFT and then select the same number of existing sheet tabs of the worksheets that you want to insert in the open workbook. For example, if you want to add three new worksheets, select three sheet tabs of existing worksheets. On the **Home** tab, in the **Cells** group, click **Insert**, and then click **Insert Sheet**.

You can also right click the selected spreadsheet tabs and then click **Insert**. On the **General** tab, click **Worksheet** and then click **OK**.

To move to the next or previous sheet tab, press CTRL+Pg Up or CTRL+Pg Dn. On the **Home** tab, in the **Cells** group, click **Format** and then under **Organize Sheets**, click **Move or Copy Sheet**.

You can also right click a selected sheet tab and then click **Move or Copy**. In the **Move or Copy** dialog box, in the **Before sheet** list, do one of the following:

- Click the sheet before which you want to insert the moved or copied sheets.
- Click **move to end** to insert the moved or copied sheets after the last sheet in the workbook and before the **Insert Worksheet** tab.

To copy the sheets instead of moving them in the **Move or Copy** dialog box, select the **Create a copy** check box. When you create a copy of the worksheet, the worksheet is duplicated in the workbook and the sheet name indicates that it is a copy, for example the first copy that you make of Sheet1 is named Sheet1 (2).

To rename the active sheet, one of the following steps can be performed:

- On the **Format** menu, point to **Sheet** and then click **Rename**.

- On the **Sheet tab** bar, right click the tab you want to rename and then click **Rename**.



Type the new name over the current name.

NOTES

To delete a worksheet, follow the given steps:

- Open the workbook.
- Click on the **Delete** button on the **Cells** group of the **Home** tab.
- Click on **Delete Sheet.**

Check Your Progress

1. What is a spreadsheet?
2. What does advanced feature permit in MS Excel 2007?
3. How can you edit a worksheet in Microsoft Excel 2007?
4. What does a cell address identifies in MS Excel 2007?
5. Write the keyboard shortcuts used to move up and down one screen in MS Excel 2007?

8.4 SELECTION IN A WORKSHEET

To select contiguous columns in the worksheet, do the following task:

- Click to select the first column heading in the desired range and then drag the pointer to the last column in the desired range. Release the mouse button to select the columns.
- You can also click to select the first column heading in the desired range. Scroll the worksheet using the scroll bars at the bottom of the worksheet as needed to display the last column in the desired range.
- Hold down the SHIFT key and click on the last column heading.

To select the entire worksheet, you need to select the first column heading and then drag the pointer across all remaining column headings in the worksheet.

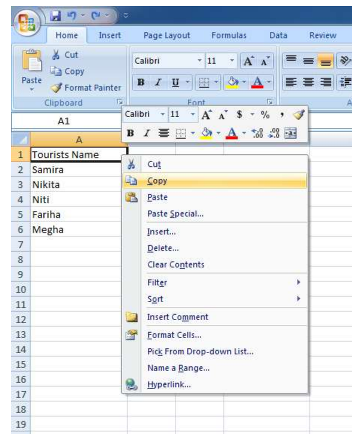
- Click to select the first column heading, scroll the worksheet until the last column is visible.
- Press the SHIFT key and click on the last column heading.

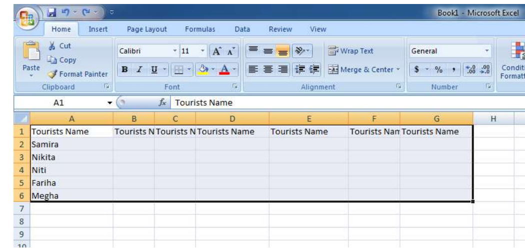
Cell Selection

To select a cell or data to be copied or cut, follow the given steps:

- Click on the **cell.**
- Click and drag the cursor to select many cells in a range.

NOTES





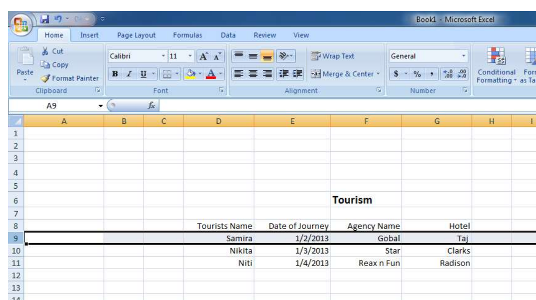
Selecting Data Range

Once you have entered your data in Microsoft Office MS Excel 2007, you should be aware of how to select cells in a worksheet. The cell cursor is made up of black border that surrounds the active cell and is called the current cell in a worksheet. In MS Excel 2007, following steps are required to select multiple cells in a worksheet:

- To select a single entire column, click a column heading, i.e., the letter or letters that indicate the column. To select multiple columns, drag across multiple column headings.
- To select a single entire row, click the row number. To select multiple rows, drag across multiple row numbers.
- To select sequential cells, click the first cell, hold down the SHIFT key and click the last cell you want. Optionally, click and drag the mouse over a group of cells to select a sequential area.
- To select non-sequential cells, click the first cell, hold down the CTRL key and click each additional cell or row or column you want to select.

Row(s)/Column(s) Selection

To select a row or column, click on the **row** or **column header**.



NOTES

Entire Worksheet Selection

To select the entire worksheet, click the small box located to the left of column A and above row 1. You can select all cells in a worksheet by pressing CTRL+A key combination

8.5 FORMULAS AND FUNCTIONS

Formulas, also called functions, are a powerful feature of MS Excel 2007. Formulas available in MS Excel 2007 are grouped into the following categories:

- **Financial:** Formulas for calculating depreciation, interest, return on an investment, loan, payment and mortgage.
- **Logical:** The logical operators of **And, True, False, If, Not** and **Or**.
- **Text:** Formulas for combining text from various cells with the concatenate formula, converting text to upper case, lower case or proper case and trimming extra spaces from text.
- **Date & Time:** Formulas for calculating the number of days between two dates, the current date and day of the week for a given date.
- **Lookup & Reference:** Formulas for horizontal or vertical lookup, a hyperlink or getting data from a Pivot Table.
- **Math & Trig:** Formulas for calculating trigonometric functions, logs, random numbers, Roman numerals, rounding and truncating.
- **Statistical:** Formulas for calculating averages, frequency, percentile, quartile and standard deviation.
- **Engineering:** Formulas for Bessel functions and conversions between numbering systems, such as octal to binary.
- **Cube:** Formulas for working with three dimensional sets.
- **Information:** Includes formulas which return a true or a false response in determining if a cell is blank or if the content of the cell is a number or text.

NOTES

8.5.1 Cell Address

In MS Excel 2007 spreadsheet, the purpose of the cell address is to identify the position of the cell. The cell address is formed by the combination of column letter and row number of a cell, such as C4 or D8. Therefore, the column letter followed by the row number can identify a cell number. Active cell can also be known as current cell. The active cell is covered with a black border and data can be entered only in the active cell. In MS Excel 2007, the cell location of cell or group of cells is indicated by a cell reference. A cell reference consists of a column letter and row number that intersects at the cell's location. The cell reference of any active cell is displayed by the Name Box. Cell references are used in formulas, functions, charts and other MS Excel 2007 commands.

8.5.2 Operators: Arithmetic, Logical, Relational, String and Reference

Operators specify the type of calculation that you want to perform on the elements of a formula. There is a default order in which calculations occur, but you can change this order by using parentheses. There are four different types of calculation operators which are named as arithmetic, comparison, text concatenation and reference. To perform basic mathematical operations, such as addition, subtraction or multiplication and to combine numbers and to produce numeric results, use the arithmetic operators summarized as follows:

Arithmetic Operator	*Example*
+ (plus sign)	Addition (3+3)
– (minus sign)	Subtraction (3–1)Negation (–1)
* (asterisk)	Multiplication (3*3)
/ (forward slash)	Division (3/3)
% (percent sign)	Percent (20%)
^ (caret)	Exponentiation (3^2)

You can compare two values using the following operators. When two values are compared by using these operators, the result is a logical value either TRUE or FALSE as follows:

Relational Operator	*Example*
= (equal sign)	Equal to (A1=B1)
> (greater than sign)	Greater than (A1>B1)
< (less than sign)	Less than (A1<B1)
>= (greater than or equal to sign)	Greater than or equal to (A1>=B1)
<= (less than or equal to sign)	Less than or equal to (A1<=B1)
<> (not equal to sign)	Not equal to (A1<>B1)

Use the string operator ampersand (&) to join or concatenate, one or more text strings to produce a single piece of text. The features of string operator is summarized as follows:

String Operator	*Example*
& (ampersand)	It concatenates two values to produce one continuous text value, such as 'Information' & 'Technology' produces the result as Information Technology.

NOTES

Reference operators are used to combine ranges of cells for calculations. The following operators are the features of reference:

Reference Operator	*Example*
: (colon)	Range operator which produces one reference to all the cells between two references including the two references (B5:B15).
, (comma)	Union operator which combines multiple references into one reference (SUM(B5:B15,D5:D15)).
(space)	Intersection operator which produces on reference to cells common to the two references (B7:D7 C6:C8).

Operator Precedence

In Microsoft MS Excel 2007, you can combine several operators in a single formula to perform the operations in the order shown in the following table. If a formula contains operators with the same precedence, for example, if a formula contains both a multiplication and division operator then MS Excel 2007 evaluates the operators from left to right.

Operator	Description
: (colon) (single space) , (comma)	Reference operators
–	Negation (as in –1)
%	Percent
^	Exponentiation
* and /	Multiplication and Division
+ and –	Addition and Subtraction
&	Connects two strings of text (concatenation)
= < > <= >= <>	Comparison

Use of Parentheses

To change the order of evaluation, enclose in parentheses the part of the formula to be calculated first. For example, the following formula produces 11 because MS Excel 2007 calculates multiplication before addition. The formula multiplies 2 by 3 and then adds 5 to the result.

=5+2*3

NOTES

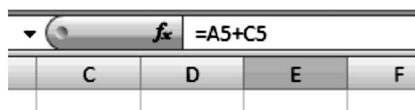
On the contrary, you can use parentheses to change the syntax as =(5+2)*3. Now the order of evaluation is changed and MS Excel 2007 will first add 5 and 2 and then multiplies the result by 3 to produce 21. In the example below, the cell addresses are used instead of numbers. The parentheses around the first part of the formula force MS Excel 2007 to calculate B4+25 first and then divide the result by the sum of the values in cells D5, E5 and F5.

=(B4+25)/SUM(D5:F5)

8.5.3 Writing Simple Formulas

A formula is referred as an equation which performs operations on a worksheet data. In MS Excel 2007, the use of formulas is to basically perform mathematical operations, namely, addition, subtraction and multiplication. They also have the ability to compare the worksheet values, taking out the average of a student's test result, join text, etc. Formulas can refer to other cells on the same worksheet, cells on other worksheets in the same workbook or cells on worksheets in other workbooks. The MS Excel 2007 can automatically recalculate the answer in case you change the data in your spreadsheet, without even re-entering or changing the formula.

A basic formula format starts with an equals sign (=) which is followed by one or more operands, which are separated by one or more operators. Operands can be values, text, cell references, ranges, defined names or function names. The various symbols that represent the arithmetic and comparison operations are known as operands.



To enter a formula, follows the given steps:

- Place the cursor in the cell where the formula will appear, i.e., E5.
- Enter an = sign. All MS Excel 2007 formulas start with the equal sign.
- Enter the expression that will produce the result you want. This can consist of operands, values, variables, and symbols which represent mathematical procedures, such as + or – to add and subtract respectively, for example A5+C5.
- When the formula is complete press ENTER key. The result of the formula will be calculated and displayed in the cell E5.
- You can see the formula in the Formula bar at the top of the screen by placing the cell pointer on the cell E5.

If there is an error in a formula, an error message is displayed which will begin with a # sign.

NOTES

MS Excel 2007 Formula Error Messages

Sometimes the formulas are not written correctly. It happens unknowingly while writing formulas. The following are some common error messages:

Error	Meaning
#####	The contents of the cell cannot be displayed correctly as the cell column is too narrow.
#REF!	Indicates that a cell reference is invalid. This message is displayed when you delete cells which involved a formula.
#NAME?	MS Excel 2007 cannot recognize text contained within a formula.

8.5.4 Copying Formula

One of the big advantages of a spreadsheet is to copy a formula or a text as often as you need. It is easier and faster to copy a formula than to rewrite it every time you need it. When you copy a formula by copying and pasting it you must be aware of cell references, whether they are absolute or relative. When you move a formula, the cell references within the formula do not change no matter what type of cell reference that you use. When you copy a formula, the cell references may change based on the type of cell reference that you use.

Select the cell that contains the formula that you want to copy. On the **Home** tab, in the **Clipboard** group, click **Copy**. Do one of the following steps:

- To paste the formula and any formatting, on the Home tab, in the Clipboard group, click **Paste**.
- To paste the formula only, on the **Home** tab, in the **Clipboard** group, click **Paste**, click **Paste Special** and then click **Formulas**.

You can paste only the formula results. On the **Home** tab, in the **Clipboard** group, click **Paste**, click **Paste Special** and then click **Values**. Verify that the cell references in the formula produce the result that you want. If necessary, switch the type of reference by doing the following:

- Select the cell that contains the formula.
- In the formula bar, you can enter or edit values or formulas in cells. It displays the constant value or formula stored in the active cell, select the reference that you want to change.
- Press F4 to switch between the combinations.

8.5.5 Cell Referencing

As already discussed, each box on the MS Excel 2007 screen is a cell and each cell can be located in a spreadsheet by means of its reference termed as cell reference. Cell on other worksheets in the same workbook and to other workbook can also be referred by cell references. References to cells in other workbooks are called links or external references. When cell references are used in formulas,

NOTES

MS Excel 2007 will calculate the answer using the data located in the referenced cells. If that data should later change, MS Excel 2007 will automatically recalculate the formula and update the answer.

Calling the cells by just their column and row labels, such as 'A1' is called **relative referencing**. When a formula contains relative referencing you copy it from one cell to another, MS Excel 2007 will not create an exact copy of the formula. It will only change the cell addresses relative to the row and column to which it is moved. For example, for a simple addition the formula in cell D1 =(A1+B1) will be copied to cell D2 as =(A2+B2). It will reflect the new row. This is called **absolute referencing** and is accomplished by placing dollar signs '$' within the cell addresses in the formula. Now, the formula in cell D1 will look like =($A$1+$B$1). **Mixed referencing** can also be used when only the row and column is to be fixed. For example, in the formula =(A$1+$B2) the row of cell A1 is fixed and the column of cell B2 is fixed.

Relative Referencing, Absolute Referencing and Mixed Referencing

A reference is a connection to something, that is to say when you type in the formula =SUM (A1;B1) you refer to the sum of the content of A1 and the content of B1. The various types of references are discussed below:

Relative Reference: Reference to rows and columns changes when we copy the formula in another cell. The formula adapts to its environment because references relate to the distance between the formula and the cells that form the formula. This is the option MS Excel 2007 offers by default. Consider the following example:

	A	B
1	15	20
2	= A1 + 2	30
3		

If we copy cell A2 in B3, because you can move one column to the right and one row down, the formula will change to: =B2+2. What varies is the reference to the cell A1 because you have copied the formula in a cell that is in the column to the right, the name of the column will change from A to B and because we have gone one row down, 1 will become 2, the result will be =B2+2. This maintains the formula that adds two to the content of the cell above.

Absolute Reference: References to rows and columns do not change when you copy the formula to another cell, the formula's reference to cells are fixed.

	A	B
1	15	20
2	= $A$1 + 2	30
3		

If you copy now the cell A2 in B3 even though you have copied it one column to the right and one row down, the formula will not change. The sign $ before the column and the row indicates that. So, B3 will show =$A$1+2.

An important difference in MS Excel 2007 spreadsheets is between absolute cell references and relative cell references. Open the MS Excel 2007 Spreadsheet as follows:

B2 *fx* = A1 + A2

	A	B	C	D	E
1	20				
2	25	45			
3					
4					

In cell B2, you need the following formula:

= A1 + A2

- Click inside cell B2 to highlight it.
- Click on cell B2 with your right mouse button, and select Copy from the menu that appears.
- Now click into cell B3.
- Again, right click the cell to get the menu. But this time click Paste option.

Your spreadsheet should now look in the following way:

	A	B	C
1	20		
2	25	45	
3		25	
4			

With cell B3 still highlighted, look at the formula bar at the top of MS Excel 2007. You should see this formula:

= A2 + A3

Click into B2 and the formula as follows:

= A1 + A2

The problem is due to cell referencing. When you clicked Copy from the menu, MS Excel 2007 does not only copy the formula. It took at look at where the cells were in the formula, relative to the B2 cell and copied this as well. From B2, the first cell reference (A1) is up one row, and left 1 column (the arrow in screen).

The second cell reference (A2) is one column to the left of cell B2.

B2 *fx* = A1 + A2

	A	B	C	D	E
1	20				
2	25	45			
3					
4					

NOTES

When you clicked into cell B3 and selected Paste from the menu, MS Excel 2007 was not only pasting the formula, it was pasting this 'up 1, left 1'. Start working at cell B3. The first cell reference appears as shown in the screen:

B2 = A1 + A2

	A	B	C	D	E
1	20				
2	25	45			
3					
4					

The second cell reference appears as shown in the screen.

	A	B	C
1	20		
2	25	45	
3		25	
4			

Thus, the first arrow is pointing to cell A2 and the second arrow is point to cell A3. This is what has been copied. MS Excel 2007 used the following formula to perform this:

= A2 + A3

But, it will be appeared as follows:

= A1 + A2

If you want the correct answer in cell B3, then you have stop MS Excel 2007 from using this relative cell referencing that it is currently doing.

Absolute cell referencing is placing a dollar symbol ($) before each letter and number. Click inside of cell B2 on your spreadsheet and change the formula as follows:

= $A$1 + $A$2

Now copy and paste it over to cell B3 again. You will get the correct answer by doing so.

B3 = $A$1 + $A$2

	A	B	C	D	E
1	20				
2	25	45			
3		45			
4					

MS Excel 2007 will use absolute formula in its calculation.

- If you need to copy and paste formulas use absolute cell references.
- Absolute referencing means typing a dollar symbol before the numbers and letters of each cell reference. You can mix absolute and relative cell references also.

NOTES

Mixed Reference: This is a combination of both references. You may have relative references for the rows and absolute for the columns or vice versa. Consider the example given below:

	A	B
1	15	20
2	= $A1 + 2	30
3		

If you copy the cell A2 in B3, the sign $ before the column will mean that the formula will not change although you move to the right. But not having the sign before the row when you copy the formula one row down it will change to 2 instead of 1 and the result will be =$A2+2. See the example given below:

	A	B
1	15	20
2	= A$1 + 2	30
3		

8.5.6 Using Functions in Formulas

A function computes the values provided and shows you the desired results, for example, a simple function **SUM()** automatically adds up all the values provided either in a range or as individual values. Some functions can even check values and take actions based on what they find. For example, by using **IF()** function, you can check whether the current value in a particular cell is greater or less than 100 and based upon the answer it gets, you can ask the function to multiply or divide this value by a different amount. Some functions can convert numbers or text to another measure or format, for example, the function **LOWER()** converts the text entry in another cell into lower case.

Typically, a function consists of two parts—function name and arguments. Arguments (or values) may in some cases be mandatory and in some cases optional. Even within a function, one of the arguments may be mandatory and another optional.

A function is a built-in formula in MS Excel 2007. A function has a name and arguments (the mathematical function) in parentheses. The following is a list of common functions used in MS Excel 2007.

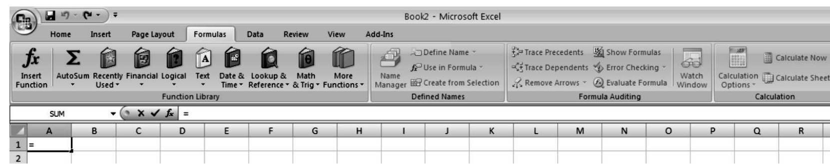
- **Sum():** Adds all cells in the argument.
- **Average():** Calculates the average of the cells in the argument.
- **Min():** Finds the minimum value.
- **Max():** Finds the maximum value.
- **Count():** Finds the number of cells that contain a numerical value within a range of the argument.

To calculate a function, follow the given steps:

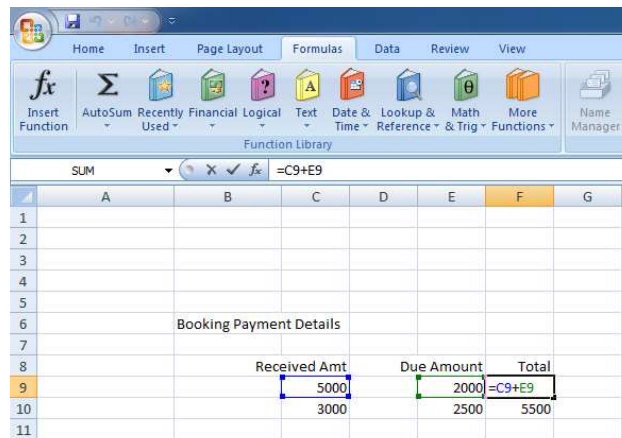
- Click the **cell** where you want to apply the function.
- Click the **Insert Function** button from Formulas tab.

NOTES

- Choose the function.
- Click **OK**.



- Write the address of first cell in the range that you want to calculate in the Number 1 text box.
- Write the address of last cell in the range that you want to calculate in the Number 2 text box.
- Click **OK**.



Mathematical Functions

In MS Excel's 2007 mathematical functions can be used to perform common mathematical operations, such as addition (SUM() function), multiplication (product function) and rounding numbers (ROUND() function). Other mathematical functions are discussed as follows:

SQRT()

In MS Excel 2007, the **SQRT()** function returns the square root of a given number. The syntax for the **SQRT()** function is as follows:

=SQRT(Number)

In the above statement, number is a positive number that you wish to return the square root for. If a negative number is entered in the number parameter, the SQRT() function will return the #NUM! error.

ROUND()

The ROUND() function in MS Excel 2007 can be used when rounding integers and decimal numbers to make them easier to work within your spreadsheets. The syntax for the ROUND() function is as follows:

=ROUND(Number, Num_digits)

NOTES

SUM()

The SUM() function in MS Excel 2007 provides a quick way to add numbers together in an MS Excel 2007. The syntax for the SUM() function is as follows:

=SUM (Number1, Number2, ..., Number255)

Up to 255 numbers can be entered into the function.

FACT()

In MS Excel 2007, the FACT() function returns the factorial of a given number. The syntax for the FACT() function is as follows:

=FACT() (Number)

In the above statement, number is a numeric value.

INT()

The INT() function is used to round a number downwards towards the next lowest number. INT() is similar to the ROUNDDOWN() function except that it always rounds a number down to the nearest whole number completely removing the decimal portion. The ROUNDDOWN() function will round a number down to a desired number of decimal places. The syntax for the INT function is as follows:

= INT (Number)

In the above statement, Number is the value to be rounded.

MOD()

The MOD() function can be used to divide numbers in MS Excel 2007. Unlike regular division, however, the MOD() function only gives you the remainder as an answer. The syntax for the MOD() function is as follows:

= MOD() (Number, Divisor)

PI()

In MS Excel 2007, the PI() function returns the mathematical constant which is equal to 3.14159265358979. The syntax for the **PI()** function is as follows:

=PI( )

POWER()

Although you can use the caret (^) operator to build a formula that raises a number to any power, the POWER() function() accomplishes the same thing. For example, to build a formula that raises 5.9 to the third power, i.e., cubes the number you can use the exponentiation operator as follows:

=5.9^3

You can have MS Excel 2007 perform the same calculation with the POWER() function by entering this formula:

=POWER(5.9,3)

In either case, MS Excel 2007 returns the same result as 205.379.

PRODUCT()

NOTES

The PRODUCT() function provides a quick way to multiple numbers in a MS Excel 2007. The advantage of using this function becomes apparent if you have several numbers to multiply together. It is easier than building a long formula. The syntax for the PRODUCT() function is as follows:

=PRODUCT(Number1, Number2, ... Number255)

In the above statement, up to 255 numbers can be entered into the function.

Date Functions

There are a number of date functions available in MS Excel 2007. Depending on your needs, you can use a date function in MS Excel 2007 to return the current date, the current time or the day of the week. Like all functions in MS Excel 2007, entering date functions is very straightforward. Just type in the function in the cell where you want the date or time to appear. The various date functions are discussed below:

NOW()

The NOW() function, one of MS Excel 2007's date and time functions, is used to add the current time and date to a spreadsheet. The syntax for the NOW() function is as follows:

= NOW ( )

The NOW() function takes no arguments.

DATE()

The DATE() function, one of MS Excel 2007's date and time functions has several uses for adding dates to a spreadsheet. It can be used to combine date elements from different locations. It can also be used to convert dates to the computer's serial date. It can be used to ensure that imported dates are formatted correctly in a spreadsheet, such as a date or a number instead of text. The syntax for the DATE() function is as follows:

= DATE(Year, Month, Day)

TODAY()

The TODAY() function, one of the MS Excel 2007's date and time functions, is used to add today's date to a spreadsheet. The syntax for the TODAY() function is as follows:

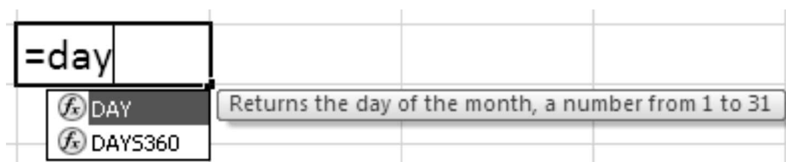
=TODAY( )

The TODAY() function takes no arguments.

NOTES

DAY()

The DAY() function returns the day of the month.



MONTH()

The function MONTH() takes one argument as MONTH(start_date) where start_date is the date for which you are trying to find the month. For example,

=MONTH(DATE(2011,3,12)

Above statement will return the value 3.

YEAR()

In MS Excel 2007, the YEAR() function returns a four digit year (a number from 1900 to 9999) given a date value. The syntax for the YEAR() function is as follows:

YEAR() (Date_value)

In the above statement, date_value is a valid date.

Logical Functions—IF(), AND(), OR(), NOT()

MS Excel 2007 uses seven logical functions, such as AND(), FALSE(), IF(), IFERROR(), NOT(), OR() and TRUE() which appear on the Logical command button's drop-down menu on the Formulas tab of the Ribbon. All the logical functions return either the logical TRUE or logical FALSE when their functions are evaluated. Here are the names of the logical functions along with their argument syntax:

- AND(logical1,logical2,...) tests whether the logical arguments are TRUE or FALSE. If they are all TRUE, the AND function returns TRUE to the cell. If any are FALSE, the AND function returns FALSE.
- IF(logical_test,value_if_true,value_if_false) tests whether the logical_test expression is TRUE or FALSE. If TRUE, the IF function returns the value_if_true argument. If FALSE, the IF function returns the value_if_false argument.
- IFERROR(value,value_if_error) tests whether the value expression is an error. IFERROR returns value_if_error if the expression is an error or value of the expression if it is not an error.
- NOT(logical) tests whether the logical argument is TRUE or FALSE. If TRUE, the NOT function returns FALSE. If FALSE, the NOT function returns TRUE.
- OR(logical1,logical2,...) tests whether the logical arguments are TRUE or FALSE. If any are TRUE, the OR function returns TRUE. If all are FALSE, the OR function returns FALSE.

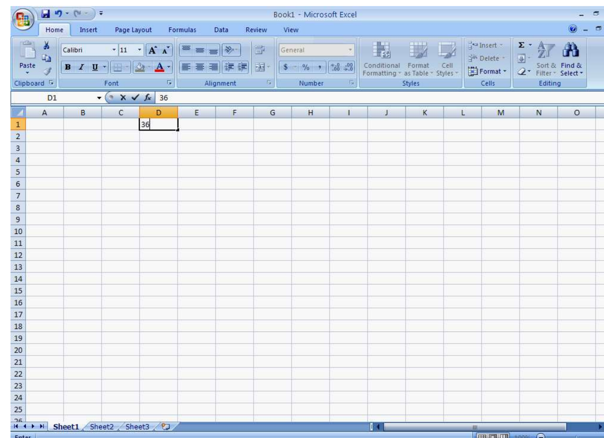
NOTES

- FALSE() takes no argument and simply enters logical FALSE in its cell.
- TRUE() takes no argument and simply enters logical TRUE in its cell.

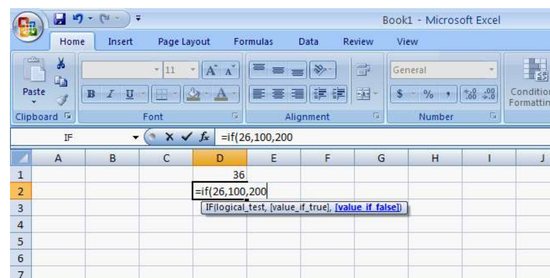
IF() Function

The IF() function is used to test whether a certain condition is TRUE or FALSE. Following steps help to get the result of IF() function:

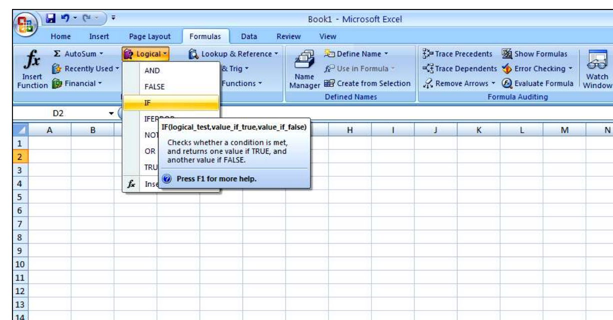
- Enter 36 into cell D1.



- Select D2 cell and type the statement as '=IF(26,100,200)'.

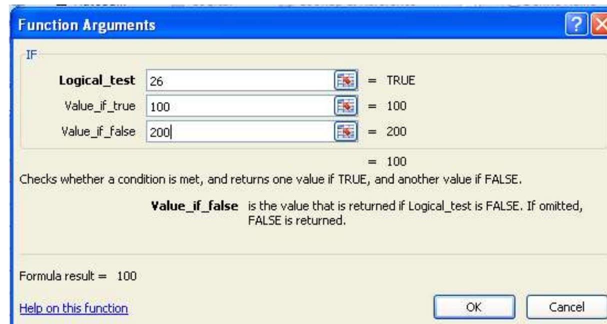


- After pressing ENTER key, value 100 will appear on D2 because 26 is less than 36 therefore 'value_if_true' condition will be executed.
- Changing the IF() function's results. Click on cell F5 which is the location of the result. Click on the Formulas tab. Choose **Logical** from the Ribbon to open the drop-down list as shown in screen. Click on IF in the list to bring up the function's dialog box.

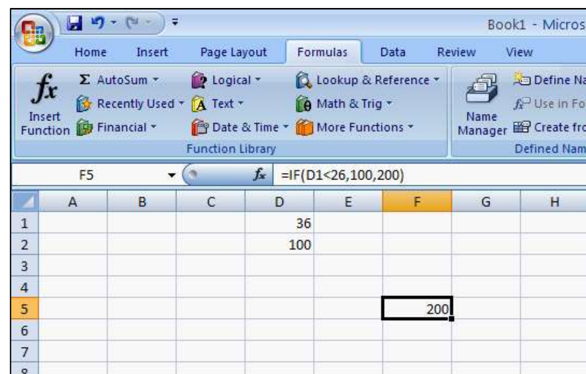


NOTES

- On the Logical_test line in the dialog box, type 26. On the Value_if_true line of the dialog box, type 100. On the Value_if_false line of the dialog box, type 200.

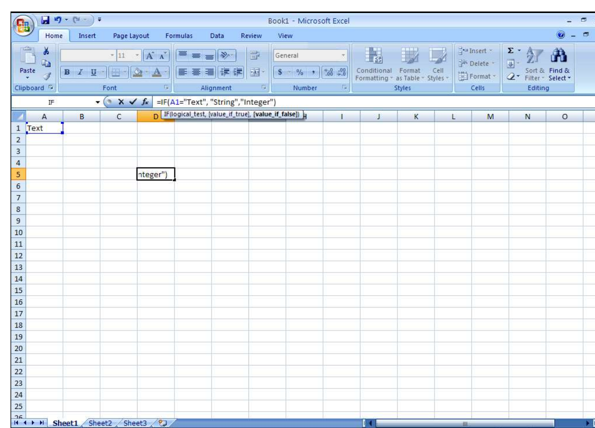


- If you click on cell F5, the complete function '= IF(D1<26,100,200)' appears in the formula bar as shown below:



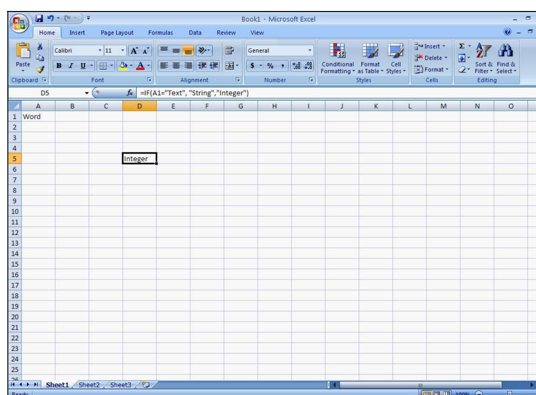
Let us take another example to explain the concept of IF() function.

Type a word "Text" in A1 cell. Keep cursor on cell D5. Type the formula =IF(A1="Text", "String", "Integer") in cell D5. Press ENTER key. Word 'String' will appear in resulted cell D5.



- If you change the word 'Word' instead of 'Text' then the value of resulted cell will contain 'Integer' as shown in screen.

NOTES



Nested IF() Function

A nested IF() function is worked when a second IF() function is placed inside the first in order to test additional conditions. It is possible to nest multiple IF() functions within one MS Excel 2007 formula. You can nest up to seven IF() functions to create a complex IF() statement. The syntax for the nested IF() function is as follows:

IF(condition1, value_if_true1, IF(condition2, value_if_true2, value_if_false2))

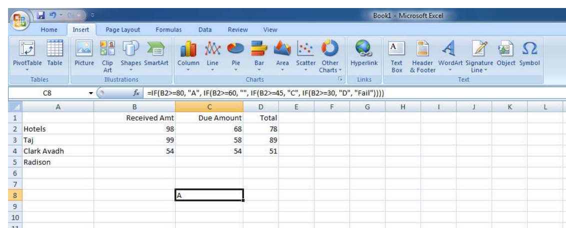
Above statement displays how to nest two IF() functions in which 'condition1' or 'condition2' is the value that you want to test, 'value_if_true' is the value that is returned if condition evaluates to **True** and 'value_if_false' returns if condition evaluates to **False**. Consider an example where using the nested IF() function the following grades will be displayed:

- A-If the student scores 80 or above.
- B-If the student scores 60 to 79.
- C-If the student scores 45 to 59.
- D-If the student scores 30 to 44.
- FAIL-If the student scores below 30.

Use the following nested IF() function to compute and display the result:

=IF(B2>=80, "A", IF(B2>=60, "B", IF(B2>=45, "C", IF(B2 >=30, "D", "Fail"))))

In the above statement, if the answer to the first question is False then control will go to next IF statement. If it is True then it will display Grade 'A' as shown in result of 'Student Examination Grade' spreadsheet.



The approach of nesting IF() function increases the flexibility of the function by a number of possible results. For example, the deductions from an employee's income will follow the approach 'the higher the income, the higher the deduction rate'.

Text Functions

Following are the text functions in MS Excel 2007.

LOWER()

In MS Excel 2007, the LOWER() function converts all letters in the specified string to lowercase.

If there are characters in the string that are not letters, they are unaffected by this function. The syntax for the LOWER() function is as follows:

=LOWER(Text)

In the above statement, text is the string to convert to lowercase.

UPPER()

In MS Excel 2007, the UPPER() function allows you to convert text to all uppercase. The syntax for the UPPER function is as follows:

=UPPER(Text)

In the above statement, text is the string that you wish to convert to uppercase.

PROPER()

In MS Excel 2007, the PROPER() function sets the first character in each word to uppercase and the rest to lowercase. The syntax for the PROPER() function is as follows:

=PROPER(Text)

In the above statement, text is the string argument whose first character in each word will be converted to uppercase and all remaining characters converted to lowercase.

LEN()

In MS Excel 2007, the LEN() function returns the length of the specified string. The syntax for the LEN() function is as follows:

=LEN(Text)

In the above statement, text is the string to return the length for.

LEFT()

In MS Excel 2007, the LEFT() function allows you to extract a substring from a string, starting from the left most character. The syntax for the LEFT() function is as follows:

=LEFT(Text, number_of_characters)

In the above statement, text is the string that you wish to extract from and the number_of_characters indicates the number of characters that you wish to extract starting from the left most character.

RIGHT()

In MS Excel 2007, the RIGHT() function extracts a substring from a string starting from the right most character. The syntax for the RIGHT() function is as follows:

=RIGHT(Text, number_of_characters)

In the above statement, text is the string that you wish to extract from and number_of_characters indicates the number of characters that you wish to extract starting from the right most character.

MID()

In MS Excel 2007, the MID() function extracts a substring from a string (starting at any position). The syntax for the MID() function is as follows:

=MID(Text, start_position, number_of_characters)

In the above statement, text is the string that you wish to extract from and start_position indicates the position in the string that you will begin extracting from. The first position in the string is 1 and number_of_characters indicates the number of characters that you wish to extract.

REPT()

In MS Excel 2007, the REPT() function returns a repeated text value a specified number of times. The syntax for the REPT() function is as follows:

=REPT(Text, Number)

In the above statement, text is the text value to repeat and number is the number of times to repeat the text value.

TRIM()

In MS Excel 2007, the TRIM() function returns a text value with the leading and trailing spaces removed. The syntax for the TRIM() function is as follows:

=TRIM(Text)

In the above statement, text is the text value to remove the leading and trailing spaces from.

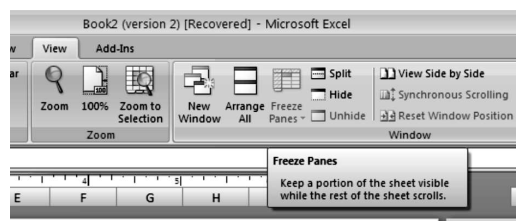
Financial Functions (PMT(), PV(), FV(), RATE(), IPMT(), NPER())

Here is the list of financial functions.

Function	*Description*
FV	This function returns the future value of an investment.
IPMT	This function returns the interest payment for an investment for a given period.
NPER	This function returns the number of periods for an investment.
PMT	This function returns the periodic payment for an annuity.
PV	This function returns the present value of an investment.
RATE	This function returns the interest rate per period of an annuity.

8.5.7 Freeze Panes

A particular portion of a worksheet can be selected to stay static while you are working on other parts of the sheet. This is done using the **Freeze Rows** and **Columns Function**. To **Freeze** a row or column, follow the given steps:



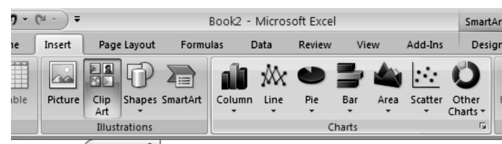
- Click on the **Freeze Panes** button on the **View** tab.
- Either select the section to be frozen or click on the defaults of top row or left column.
- To unfreeze, click on the **Freeze Panes** button.
- Click on **Unfreeze.**

Check Your Progress

6. What is a cell cursor?
7. Which key combination is used to select all cells in a worksheet?
8. Define the term cell address.
9. What does MS Excel 2007 function compute?
10. What does FACT() function return?
11. What does LEN() function return?

8.6 CHARTS

Charts help you to present the information of the worksheet in a graphic format. MS Excel 2007 provides various types of charts which includes **Column**, **Line**, **Pie**, **Bar**, **Area**, **Scatter** and **Other charts**. To view the charts click on the **Insert tab** on the Ribbon.

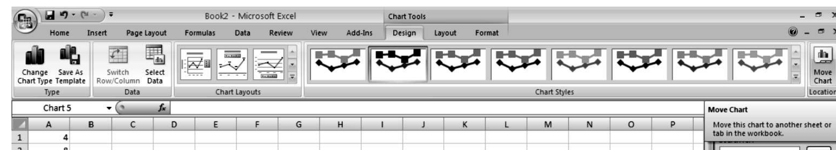


Creating a Chart: To create a chart, follow the given steps:

- Select the cells that contain the data you would like to use in the chart.
- Click on the **Insert** tab on the Ribbon.
- Click on the **Charts**, choose the type of chart you want to create.

NOTES

Moving a Chart: You can modify and move a created chart. To move the chart, follow the given steps:



- Click on the chart and drag it to the location, you want the chart to insert in the same worksheet.
- Click on the **Move Chart** button on the **Design** tab.
- Select the desired location in the same or another worksheet.

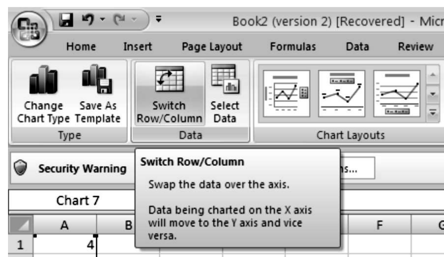
To change the data included in the chart, follow the given steps:



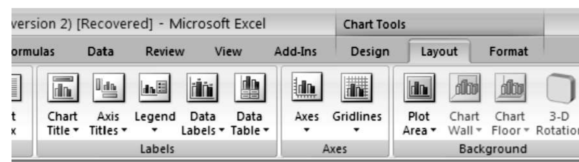
- Click on the chart.
- Click on the **Select Data** button on the **Design** tab.

To reverse the data displayed in the rows and columns, follow the given steps:

- Click on the chart.
- Click on the **Switch Row/Column** button on the **Design** tab.

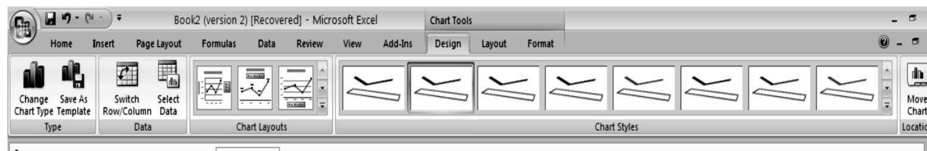


To modify the labels and titles, follow the given steps:

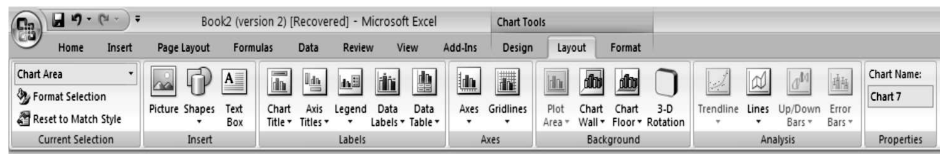


- Click on the chart.
- On the **Layout** tab, click on the **Chart Title** or the **Data Labels** button.
- Modify the title and Press ENTER key.

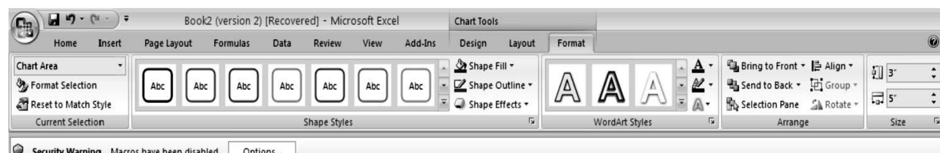
Chart Tools: The Chart Tools appear on the Ribbon when you click on the chart. The tools are located on three different tabs that are Design, Layout and Format. Within the **Design** tab you can control the chart type data, layout, styles and location.



The **Layout** tab controls inserting pictures, shapes and text boxes, labels, axes, background and analysis.

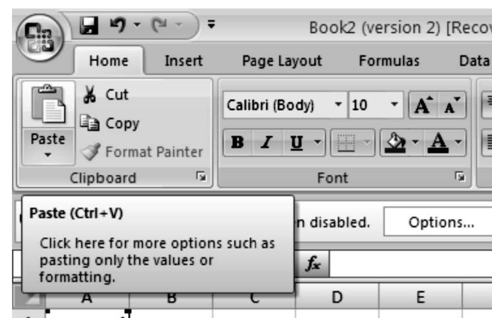


Within the **Format** tab you can modify shape styles, word styles, chart arrangement and alignment, and size of the chart.



Copy a Chart to Word

- Select the chart.
- Click on **Copy** on the **Home** tab.
- Go to the **Word** document where you want the chart to be located.
- Click on **Paste** on the **Home** tab.



Creating an Embedded Chart

An embedded chart is referred as a chart that appears right within the worksheet so that when you save or print the worksheet, you save or print the chart along with it. By default, the chart is placed on the worksheet as an embedded chart. You can also choose to create a chart in its own chart sheet in the workbook at the time you create it. Embed a chart on worksheet when you want to print the chart

along with its supporting worksheet data. Place a chart on its own sheet when you intend to print the charts of the worksheet data separately. The steps for creating an embedded chart are exactly same as for creating a simple chart.

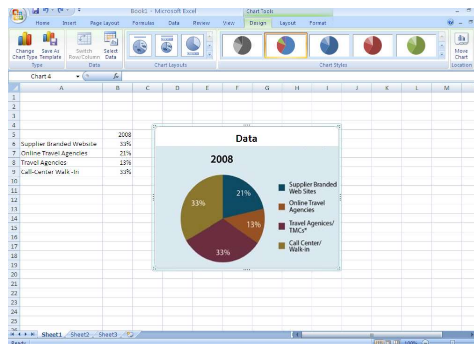
NOTES

Creating Chart Sheet

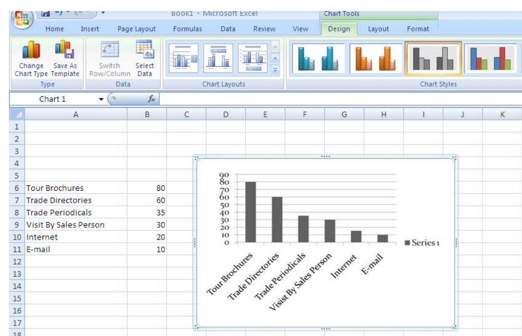
By default, when you create a chart, MS Excel 2007 embeds the chart in the active worksheet. You can move a chart to another worksheet or to a chart sheet. A chart sheet is a sheet dedicated to a particular chart. By default, MS Excel 2007 names each chart sheet sequentially starting with Chart1. You can change the name of the **chart sheet** as per your choice.

Formatting Chart: Title, Gridlines, Legends and Resizing

To add chart, you need to type suitable title for the prepared chart. Suitable title is provided to the prepared chart when it is inserted in the worksheet. Subject names can be displayed on your chart by selecting the tab marked as **Data Labels**. Chart's appearance can be viewed by seeing the previews of the changes you make when you check or uncheck the property boxes.



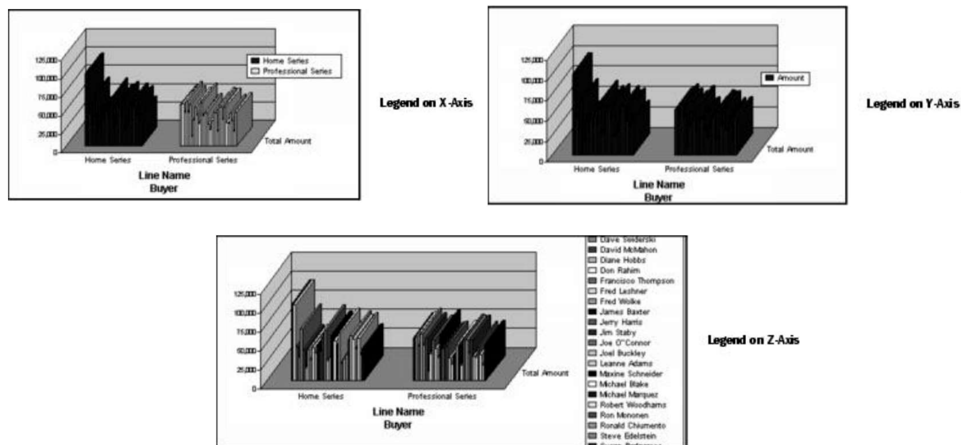
Data can be displayed in horizontal and vertical chart gridlines to make it easier to read and understand. Gridlines extend from any horizontal and vertical axes across the plot area. In a 2-D chart, the area is bounded by the axes including all data series. In a 3-D chart, the area is bounded by the axes, including the data series, category names, tick mark labels and axis titles of the chart. Depth gridlines can also be displayed in 3-D charts. Note that gridlines cannot be displayed for those kind of chart types that do not have any axes, such as pie charts and doughnut charts.



NOTES

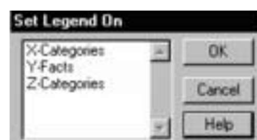
You can choose the axis along which you want to distinguish your data by setting the chart legend on that axis. This is a significant way to view values on the selected axis without rearranging the values in the Outliner. A chart legend can be set on the X-, Y- or Z-axis. You can also reposition or resize a legend to take advantage of either the horizontal or vertical space within the chart area.

The following three examples show how the legend are placed on different axes to alter the appearance and data shown by the same chart. In the first example, the legend has been set on the X (categories) Axis. In the second example, the legend has been set on the Y (facts) Axis. In the third example, the legend has been set on the Z (clusters) Axis:



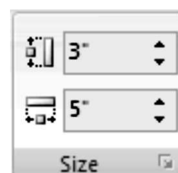
To set the axis used for a chart legend from the **Format** menu, follow the given steps:

- Choose **Format** → **Set Legend On**.
- The **Set Legend On** dialog box appears.
- Select the axis on which you want to set the legend and click **OK**.



To resize a chart, do one of the following:

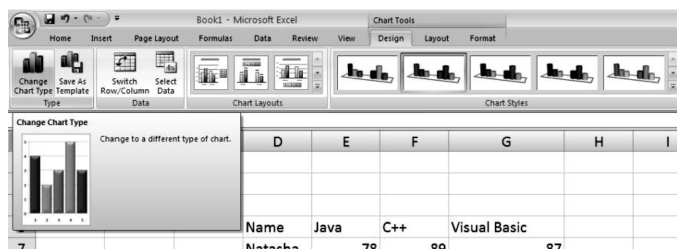
- To change the size manually, click the chart and then drag the sizing handles to the size that you want.
- To use specific height and width measurements, on the **Format** tab, in the **Size** group, enter the size in the **Shape Height** and **Shape Width** box.



NOTES

Changing Chart Types

Any change you can make to a chart that is embedded in a worksheet, you can also make to a chart sheet. For example, you can change the chart type from a column chart to a bar chart and for this following steps are required:



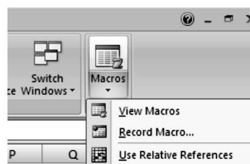
- Start Microsoft Excel 2007 and open a workbook from your documents that contains a chart or you can create a chart from data in an existing workbook.
- Right click on top of the chart that you want to change to a different type of chart and click '**Change Chart Type**'. The '**Change Chart Type**' dialog box will open in the middle of the screen.
- Click the type of chart you would like to change to on the left side of the '**Change Chart Type**' dialog box.
- Select the specific chart you would like to change your existing chart to by clicking it in the right portion of the '**Change Chart Type'** dialog box. The chart you have chosen will be highlighted.
- Click the '**OK**' button to close the '**Change Chart Type**' dialog box and to change the selected chart type to the new chart type that you have specified.

Check Your Progress

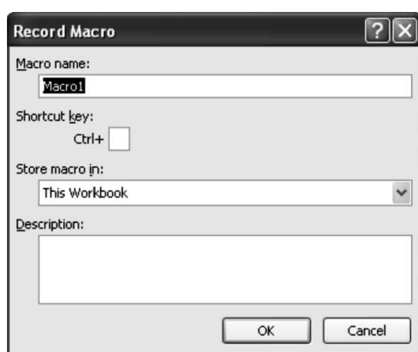
12. How do modifying fonts in MS Excel 2007 help?
13. Write the function of Layout tab.
14. Name the chart sheet which is sequentially started in MS Excel 2007.

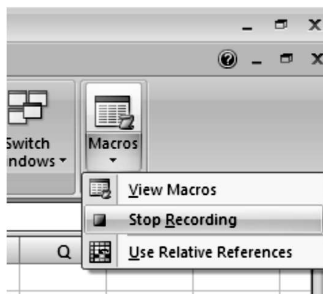
8.7 CREATING AND USING MACROS

Macros are advanced features which speed up editing or formatting in an MS Excel 2007 worksheet. They record the sequences of menu selections selected by you so that a series of actions can be successfully performed in one step.



NOTES



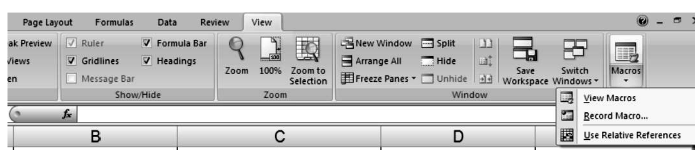


Recording a Macro: To record a Macro, follows the given steps:

- Click on the **View** tab on the Ribbon.
- Click on **Macros.**
- Click on **Record Macro...**
- Enter a name for Macro (without spaces).
- Enter a **Shortcut Key.**
- Enter a **Description** and click OK.
- Record the **Macro.**
- Click on **Marcos.**
- Click on **Stop Recording** to stop macro recording.

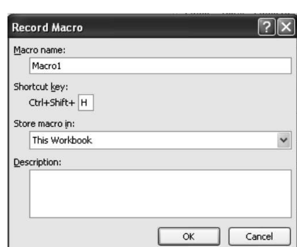
Assigning a Keyboard Shortcut to a Macro

Macro in MS Excel 2007 is used to record each command and action that is required to perform task. The created and recorded macro can further be used to carry out the similar task in a worksheet. You can run the specified Macro for performing desired action. The simplest way to create a Macro in MS Excel 2007 is to use the Macro recorder. The required steps are as follows:



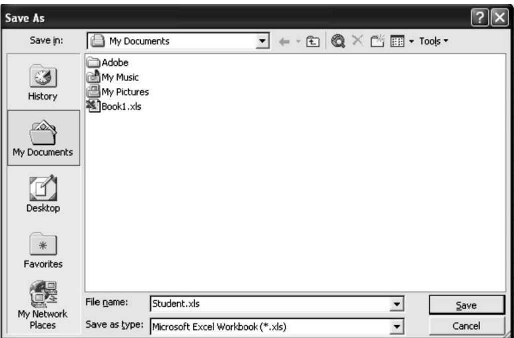
- Open the workbook where you want to use the Macro.
- Click on **View** → **Record Macro** and then select **Record New Macro** from the menu.

  You will now see the **Record Macro** dialog box as shown below:



NOTES

After setting the shortcut key for creating Macro as CTRL+SHIFT+H and Macro name as **Macro1**, the given screen will appear which shows the recording Macro button. This button records all the activities which will be issued by user. Type the valuable data in the worksheet. Save the file name as 'Student.xls'.



Running Macro

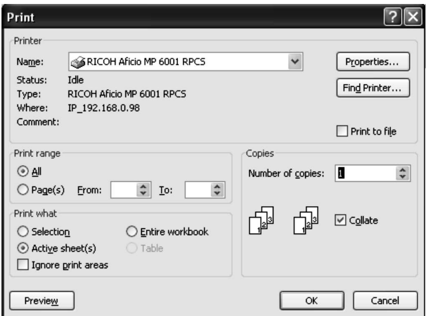
To run a Macro from the Keyboard shortcut, simply press the keys that you have assigned to run the Macro.

- Click on **Macros.**
- Click on **View Macros.**
- Select the **Macro** and click on **Run.**



8.8 PRINTING WORKSHEETS

You can choose the spreadsheet you want to print. To print a worksheet, do the following steps:



- Click **Print** on the **File** menu.
- Under **Print what**, select **Entire** workbook radio button.

You will get the **Print** tab where you can set the '**Print range**', '**Print what**', **'Number of copies:'** options as shown in the screen .

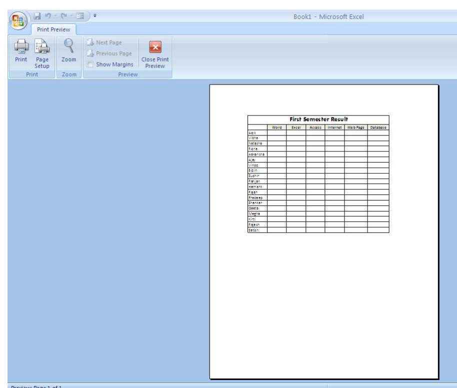
NOTES

Previewing Worksheet

To preview worksheet, go to the **Office** button in which **Print** option is given. To view the spreadsheet in Print Preview, follow the given steps:

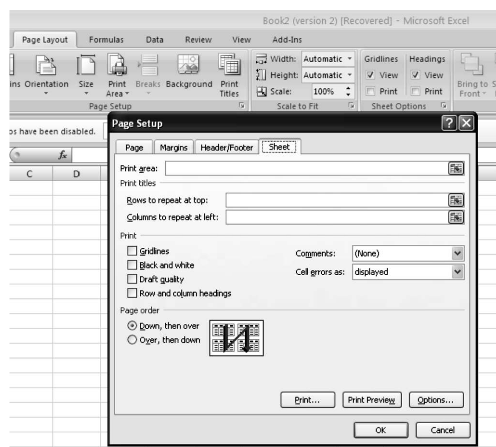
- Click on the **Office button** on the Ribbon.
- Select the arrow to the right of **Print**.
- Choose **Print Preview.**

The spreadsheet appears on screen as it will be printed. You can proceed to print the document from here or you can change options and settings to make the printed output look different. When you select **Print Preview** option the worksheet page will look as shown in the screen below:



Page Setup Margin, Paper Size and Orientation Setting

Set Print Titles: The Print Titles function permits you to repeat the column and row headings at the beginning of each new page to make reading a multiple page sheet easy to read when printed. To print titles, follow the given steps:



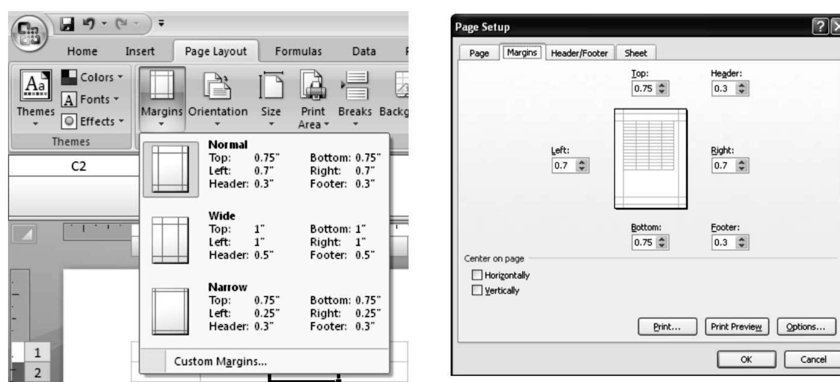
- Click on the **Page Layout** tab on the Ribbon.

- Click on the **Print Titles** button.
- In the **Print Titles** section, click on the box to select the rows/columns to be repeated.

NOTES

- Select the row or column.
- Click on the **Select Row/Column Button.**
- Click on OK.

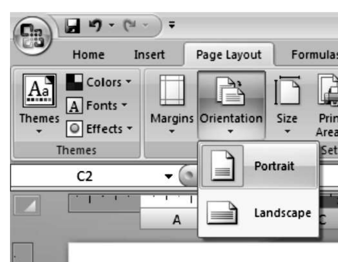
Set Page Margins: To set the page margins, follow the given steps:



- Click on the **Margins** button on the **Page Layout** tab.
- Select any one of the given options.
- Click on **Custom Margins.**
- Complete the boxes to set margins.
- Click on OK.

Changing Page Orientation: To change the page orientation from portrait to landscape, follow the given steps:

- Click on the **Orientation** button on the **Page Layout** tab.
- Select **Portrait** or **Landscape.**



Printing an Area of a Worksheet

A print area can be defined for specific sections on the worksheet that are printed frequently. A print area can be defined as one or more selected range of cells that are selected for printing so that the entire worksheet does not need to be printed. Therefore, when print command is executed after selected a print area, only the selected cells are printed. If required, cells can be added to enlarge the print area. For printing the entire worksheet, the print area is simply needed to be cleared.

NOTES

To set a print area, follow the given steps:

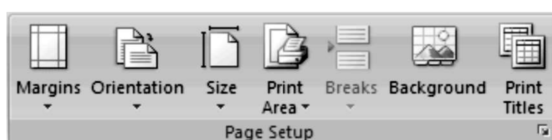
- On the worksheet, select the cells that you want to define as the print area.
- On the **Page Layout** tab, in the **Page Setup** group, click **Print Area** and then click **Set Print Area**.

The print area that you set is saved when you save the workbook. You can add cells to an existing print area by performing the following steps:

- On the worksheet, select the cells that you want to add to the existing print area.
- On the **Page Layout** tab, in the **Page Setup** group, click **Print Area** and then click **Add to Print Area**.

A print area can be cleared by performing the following steps:

- Click anywhere on the worksheet for which you want to clear the print area.
- On the **Page Layout** tab, in the **Page Setup** group, click **Clear Print Area**.



Printing Worksheet with/without Gridlines

To make a printed worksheet or workbook easier to read, you can print the worksheet or workbook with gridlines displayed around the cells.

- Select the worksheet or worksheets that you want to print.
- When multiple worksheets are selected, **Group** menu appears in the title bar at the top of the worksheet. To cancel a selection of multiple worksheets in a workbook, click any unselected worksheet. If no unselected sheet is visible, right click the tab of a selected sheet and then click **Ungroup Sheets** on the shortcut menu.
- On the **Page Layout** tab, in the **Sheet Options** group, select the **Print** check box under **Gridlines**.
- Click **Microsoft Office Button** and then click **Print**. You can also press CTRL+P which are assigned as keyboard shortcut. Worksheets print faster if you print without gridlines.

Adding Standard/Customized Header and Footer in Worksheet

To create a header or footer in a worksheet, follow the given steps:

- Click on the **Header & Footer** button on the **Insert** tab.
- It will display the **Header & Footer Design Tools** tab.

NOTES

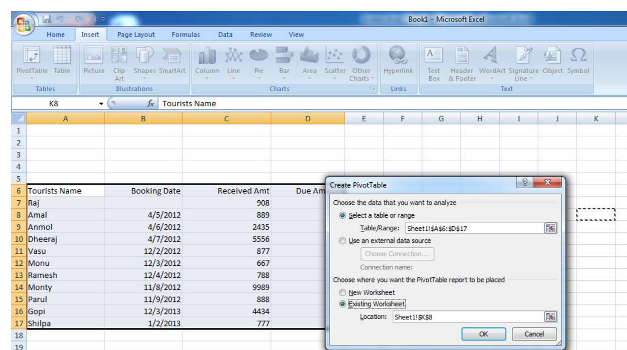
- To switch between the Header and Footer, click on the **Go to Header** or **Go to Footer** button.
- To insert text, enter the text in the header or footer.
- To enter preprogrammed data, such as page numbers, date, time, file name or sheet name, click on the respective button.
- To change the location of data, click on the desired cell.

PivotTable and PivotChart

The PivotTables are interactive tables that allow the user to group and summarize large amounts of data in a concise, tabular format for easier reporting and analysis. A PivotChart is the graphical representation of PivotTables. When you create a PivotTable or PivotChart report and can use several different types of source data. The source data refers to the list or table that is used to create a PivotTable or PivotChart report. Source data can be taken from an MS Excel 2007 list or range, an external database or another PivotTable report. A PivotTable report is used to summarize, analyse, explore and present summary data. This report is visualized in the summary data in a PivotTable report and to easily see comparisons, patterns and trends. Both a PivotTable report and a PivotChart report enable you to make informed decisions about critical data in your enterprise.

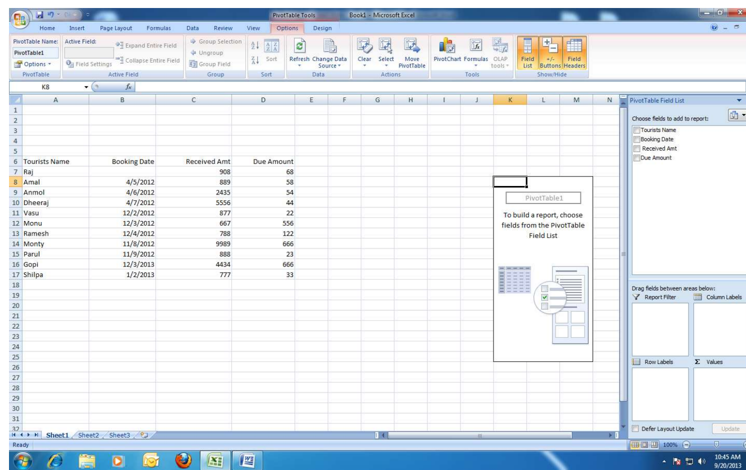
Creating a PivotTable: To create a PivotTable, following steps are required:

- Prepare a suitable data on which you want to create a PivotTable and select the range in **Table/Range:** bar as shown in the screen below.

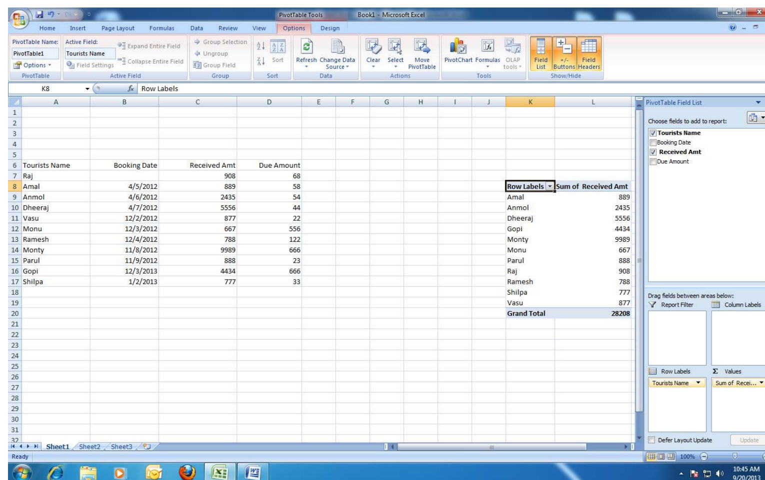


- After clicking on OK button, **'PivotTable Field List'** appears where you can choose the selected fields on which you want to prepare the PivotChart.

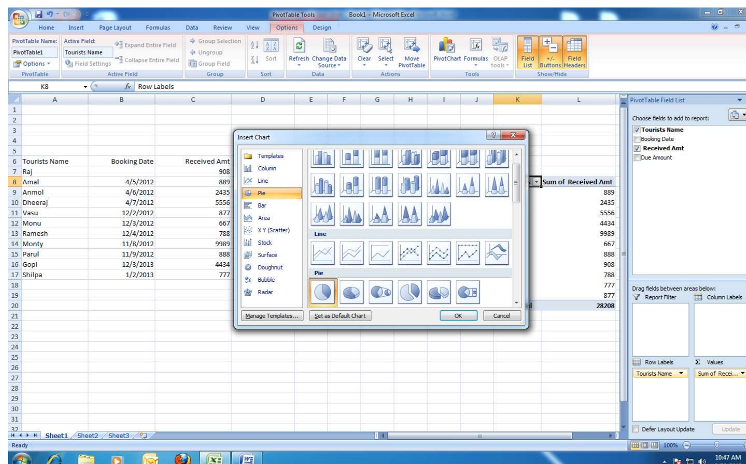
NOTES



- Once you select the desired fields, the **PivotTable** is prepared as shown in screen.

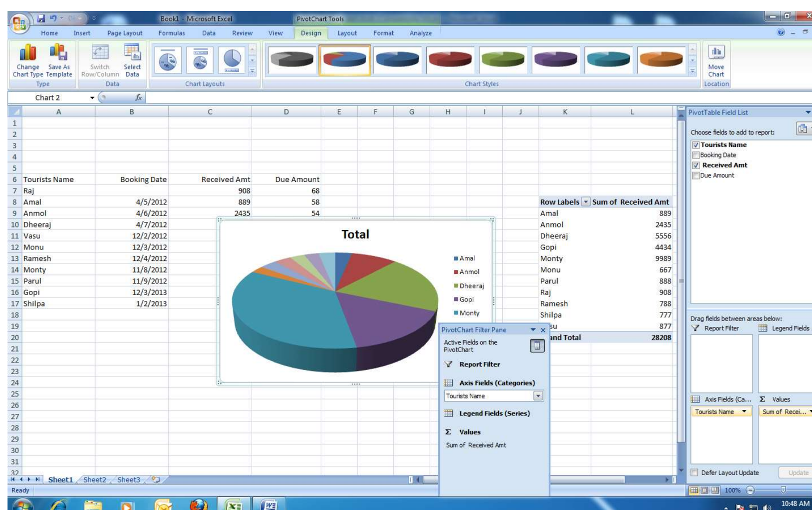


- You can create a **PivotChart** using tool. This tool will provide **'Insert Chart'** option where you can create a suitable chart to depict the selected data.



NOTES

- Once you select the chart, for example 'Pie' chart then the result appears as follows:



8.9 ESTABLISHING WORKSHEET LINKS

To summarize and report results from separate worksheets, you can consolidate data from each separate worksheet into a master worksheet. The worksheets can be in the same workbook as the master worksheet or in other workbooks. When you consolidate data, you are assembling data so that you can more easily update and aggregate it on a regular or ad hoc basis. For example, if you have a worksheet of expense figures for each of your regional offices, you might use a consolidation to roll up these figures into a corporate expense worksheet. This master worksheet might contain sales totals and averages, current inventory levels, and highest selling products for the entire enterprise. To consolidate data, use the **Consolidate** command in the **Data Tools** group on the **Data** tab. After you have consolidated data from multiple worksheets, you can change the in which the data is consolidated. For example, to add worksheets from new regional offices, delete worksheets from departments that no longer exist or change formulas with 3-D references. The term '3-D reference' is a range that spans two or more worksheets in a workbook. You change a consolidation by changing the formulas or by editing the formulas, such as changing the function or expression. Regarding cell references, you can do one of the following:

To Change a Consolidation made by Position or Category

You can change the consolidation only if you have not previously selected the **'Create links to source data'** check box in the **Consolidate** dialog box. If the check box is selected, click **Close** and then recreate the consolidation. Click the upper left cell in the consolidated data.

- On the **Data** tab, in the **Data Tools** group, click on **Consolidate**.

- The new source range must have either data in the same positions if you previously consolidated by position or column labels that match those in the other ranges in the consolidation if you previously consolidated by category.
- If the worksheet is in another workbook, click **Browse** to locate the file and then click OK to close the Browse dialog box. The file path is entered in the **Reference** box followed by an exclamation point.
- Type the name for the range and then click **Add**.
- Adjust the size or shape of a source range.
- Under **All references**, click the source range that you want to change.
- In the **Reference** box, edit the selected reference.
- Click on **Add**.

NOTES

To Delete a Source Range from the Consolidation

- Under **All references**, click the source range that you want to delete.
- Click on **Delete**.

To Make the Consolidation Update Automatically

Select the check box if the worksheet is in another workbook. Once you select this check box, you will not be able to change which cells and ranges are included in the consolidation.

- Select the **Create links to source data** check box.
- To update the consolidation with the changes, click OK.

If the data to consolidate is in different cells on different worksheets, you need to add, change or delete the cell references to other worksheets. For example, to add a reference to cell G3 in a Facilities worksheet that you have inserted in the following Marketing worksheet, you must edit the formula as shown in the screen. The MS Excel 2007 sheet appears before as follows:

	A	B	C	D	E
1					
2	=SUM(Sales!B4, HR!F5, Marketing!B9)				
3					

Sales / HR / Marketing / **Consolidation**

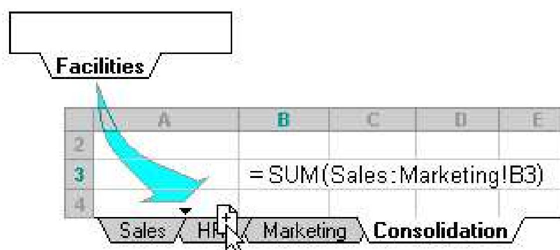
Once the multiple worksheets, such as 'Sales', 'HR', 'Marketing' and 'Facilities' are consolidated, the result appears as follows:

	A	B	C	D	E	F
1						
2	= SUM(Sales!B4, HR!F5, Marketing!B9, Facilities!G3)					
3						

Sales / HR / Marketing / Facilities / **Consolidation**

To add another worksheet to the consolidation, move the sheet into the range that your formula refers to. For example, to add a reference to cell B3 in the Facilities worksheet, move the Facilities worksheet between the 'Sales' and 'HR' sheets as shown in the screen.

NOTES



Because, the formula contains a 3-D reference to a range of worksheet names, **'Sales:Marketing!B3'** as shown in the screen, all worksheets in the range are included in the new calculation.

During a consolidation, MS Excel 2007 enables you to link the data in the source areas specified in the **'All references'** list box of the **'Consolidate'** dialog box to the destination area in the new worksheet. Thus any change that you make to the values in the destination area of the consolidation worksheet is applied to all references. To create links between the source worksheets and destination worksheet, just select the **'Create Links to Source Data'** check box to the **'Consolidate'** dialog box prior to performing the consolidation. When you perform a consolidation with linking, MS Excel 2007 creates the links between the source areas and the destination area by outlining the destination area. Each outline level created in the destination area holds rows or columns that contain the linking formulas to the consolidated data.

Check Your Progress

15. What is the simplest way to create a macro in MS Excel 2007?
16. Why a PivotTable report is used?
17. What is '3-D reference' in consolidating data?

8.10 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A spreadsheet is an electronic document which saves various types of data in columns and rows.
2. The advanced feature permits you to specify the options for editing, copying, pasting, printing as well as displaying formulas, calculations and other general settings.
3. You can edit a worksheet if any incorrect information has been entered into a cell. To do so, click on the cell and enter the correct information. Typing replaces whatever is in the cell. Editing a spreadsheet includes copy, cut, paste, move, changing column width/row height, cell alignment and formatting, font and number formatting, inserting and deleting cell(s)/row(s)/column(s), insert/copy/move/rename/delete a worksheet.

NOTES

4. In MS Excel 2007 spreadsheet, a cell address identifies the location of the cell.
5. In MS Excel 2007, to move up or down one screen the keyboard shortcuts used are Pg Up and Pg Dn keys.
6. The cell cursor is a black border that surrounds the active cell sometimes called the current cell in a worksheet.
7. CTRL+A key combination is used to select all cells in a worksheet.
8. The cell address is a combination of column letter and row number of a cell, such as C4 or D8.
9. A function computes the values provided and shows you the desired results; e.g., a simple function SUM() automatically adds up all the values provided either in a range or as individual values.
10. In MS Excel 2007, the FACT() function returns the factorial of a number.
11. In MS Excel 2007, the LEN() function returns the length of the specified string.
12. Modifying fonts in MS Excel 2007 helps you to emphasize titles and headings.
13. The Layout tab controls inserting pictures, shapes and text boxes, labels, axes, background and analysis.
14. By default, MS Excel 2007 names each chart sheet sequentially starting with Chart1.
15. The simplest way to create a macro in MS Excel 2007 is to use the macro recorder.
16. A PivotTable report is used to summarize, analyse, explore and present summary data.
17. The term '3-D reference' is a range that spans two or more worksheets in a workbook while consolidating data.

8.11 SUMMARY

- Microsoft Excel 2007 is the newest version of spreadsheet software in the Microsoft 2007 Office Suite.
- MS Excel 2007 files are referred as spreadsheets. This is a generic term, which sometimes means a workbook (file) and sometimes means a worksheet (a page within the file).
- Data files created with MS Excel 2007 are called workbooks. MS Excel 2007 files by default contain three blank worksheets. This gives you the flexibility to store related data in different locations within the same file.

NOTES

- In MS Excel 2007, the columns are lettered A to Z and then continuing with AA, AB, AC, and so on, and the rows are numbered from 1 to 1,048,576.
- MS Excel 2007 provides a wide range of customizable options that help you to create an MS Excel 2007 workbook of required specifications.
- The **Ribbon** is the panel at the top portion of the spreadsheet. It includes seven tabs namely, **Home, Insert, Page Layout, Formulas, Data, Review** and **View**.
- The **Quick Access Toolbar** can be customized as per the user need and contains commands that you use most frequently. You can place the Quick Access Toolbar above or below the Ribbon. To change the location of the Quick Access Toolbar, click on the arrow at the end of the toolbar and click **Show Below** the Ribbon.
- **Mini Toolbar** is a new feature in Microsoft Office 2007. This is a floating toolbar and is displayed when you select text or right click any text. It displays the common formatting tools, such as Bold, Italic, Fonts, Font Size and Font Color.
- The Proofing feature permits you to personalize the options for correcting words and formats of your text.
- If you hold down the CTRL key and then press the Home key, you move to the beginning of the worksheet.
- If there is an error in a formula, an error message is displayed which will begin with a # sign.
- The ROUND() function in MS Excel 2007 can be used when rounding integers and decimal numbers to make them easier to work within your spreadsheets.
- In MS Excel 2007, you can change any column width or row height in your worksheets to improve the readability and appearance of data.
- Within the Format tab you can modify shape styles, word styles and size of the chart.
- In a 3-D chart, the area is bounded by the axes, including the data series, category names, tick mark labels and axis titles of the chart.
- To run a Macro from the keyboard shortcut, simply press the keys that you have assigned to run the Macro.
- The PivotTable report is an interactive, cross tabulated MS Excel 2007 report that summarizes and analyses data, such as database records from various sources including ones that are external to MS Excel 2007.
- The PivotTable make easy to arrange and summarize complicated data based on details. When you create a PivotTable or PivotChart report, you can use any of several different types of source data. The source data refers to the list or table that is used to create a PivotTable or PivotChart report.

- To summarize and report results from separate worksheets, you can consolidate data from each separate worksheet into a master worksheet.

8.12 KEY WORDS

NOTES

- **Cell address:** The combination of a column coordinate and a row coordinate
- **Popular features:** To personalize your work environment using the mini toolbar, color schemes, default options for new workbooks, customize sort and fill sequences user name
- **Operators:** Symbols used to represent the various arithmetic and comparison operations you can perform on the operands
- **PI():** This function returns the mathematical constant which is equal to 3.14159265358979
- **NOW():** This function is used to add the current time and date to a spreadsheet
- **LOWER():** This function converts all letters in the specified string to lowercase
- **Level buttons:** Buttons that represents a level of organization in a worksheet and clicking a level button hides all levels of detail below that of the button you clicked
- **Charts:** Graphical representation of data
- **Macro buttons:** Button that records all the activities in MS Excel 2007 which will be issued by users
- **Print area:** Area refers to one or more ranges of cells that you designate to print when you do not want to print the entire worksheet

8.13 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What is Microsoft Excel 2007?
2. Write the steps to select data range.
3. Which string operator is used to concatenate the two text values?
4. Write the syntax of MOD() function.
5. Write the steps to cut and paste the formula.
6. How any chart type is changed in MS Excel 2007?
7. What are macros?
8. Why previewing worksheet is a required step before taking the printout?
9. How a PivotTable is created?

10. Why is PivotTable report used?
11. What happens when you consolidate data?

Long-Answer Questions

NOTES

1. Describe the difference between worksheet, workbook and workspace with the help of suitable examples.
2. Explain the features of Microsoft Office button in MS Excel 2007.
3. Describe the concept of selection in a worksheet.
4. Discuss the categories of formulas in MS Excel 2007.
5. Explain the syntax of text functions: LOWER(), UPPER(), PROPER(), LEN(), LEFT()RIGHT(), MID(), REPT() and TRIM().
6. Explain briefly the concept of editing worksheet.
7. Discuss the steps required in creating chart and embedded chart, moving chart and formatting chart.
8. How a keyboard shortcut is assigned in creating and executing a macro? Explain briefly.
9. Explain the steps of creating header and footer, setting page margins and printing an area of a worksheet.
10. Explain PivotTable and PivotChart with the help of suitable examples and illustrations.
11. Discuss the process to make the consolidation update automatically.

8.14 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

BLOCK - III
DATA ANALYSIS

UNIT 9 DATA ANALYSIS

NOTES

Structure

9.0 INTRODUCTION

Data analysis is a process of inspecting, cleansing, transforming and modelling data with the goal of discovering useful information, informing conclusions and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, and is used in different business, science, and social science domains. Nowadays, data analysis plays a role in making decisions more scientific and helping businesses operate more effectively. Analysts may use robust statistical measurements to solve certain analytical problems. Hypothesis testing is used when a particular hypothesis about the true state of affairs is made by the analyst and data is gathered to determine whether that state of affairs is true or false. Hypothesis testing involves considering the likelihood of Type I and Type II errors, which relate to whether the data supports accepting or rejecting the hypothesis.

Regression analysis involves identifying the relationship between a dependent variable and one or more independent variables. A model of the relationship is hypothesized, and estimates of the parameter values are used to develop an estimated regression equation. Various tests are then employed to determine if the model is satisfactory. If the model is estimated satisfactory, the estimated regression equation can be used to predict the value of the dependent variable given values for the independent variables.

NOTES

Correlation and regression analysis are related in the sense that both deal with relationships among variables. The correlation coefficient is a measure of linear association between two variables. Values of the correlation coefficient are always between -1 and +1. A correlation coefficient of +1 indicates that two variables are perfectly related in a positive linear sense, a correlation coefficient of -1 indicates that two variables are perfectly related in a negative linear sense, and a correlation coefficient of 0 indicates that there is no linear relationship between the two variables.

In this unit, you will study about the data analysis techniques, preparation of frequency distribution, calculation of mean, standard deviation, coefficient of variation, correlation coefficient, regression coefficients, trend line using method of least squares and moving averages.

9.1 OBJECTIVES

After going through this unit, you will be able to:

- Discuss the significance of data analysis
- Prepare frequency distribution graphs
- Calculate mean, standard deviation and coefficient of variation
- Explain and calculate the correlation coefficient and the regression coefficients
- Elaborate on the trend line using method of least squares and moving averages

9.2 PREPARATION OF FREQUENCY DISTRIBUTIONS

In a graph, the independent variable should always be placed on the horizontal or *x*-axis and the dependent variable on the vertical or *y*-axis.

Line Graph

Here, the points are plotted on paper (or graph paper) and joined by straight lines. Generally, continuous variables are plotted by the line graph.

Example 1: The monthly averages of Retail Price Index from 1996 to 2003 (Jan. 1996 = 100) were as follows:

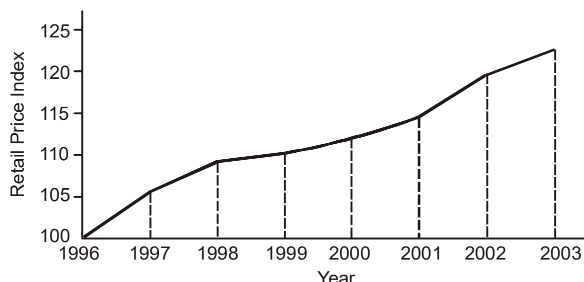
Year	1996	1997	1998	1999	2000	2001	2002	2003
Retail Price Index	100	105.8	109.0	109.6	110.7	114.5	119.3	122.3

Draw a diagram to display these figures.

Solution: Here, years are plotted along the horizontal line and the retail price index along the vertical line.

Erect perpendiculars to horizontal line from the points marked as retail price index for the years 1997, 1998, ..., 2003 and cut off these ordinates according to the given data and thus various points will be plotted on the paper. Join these points by straight lines.

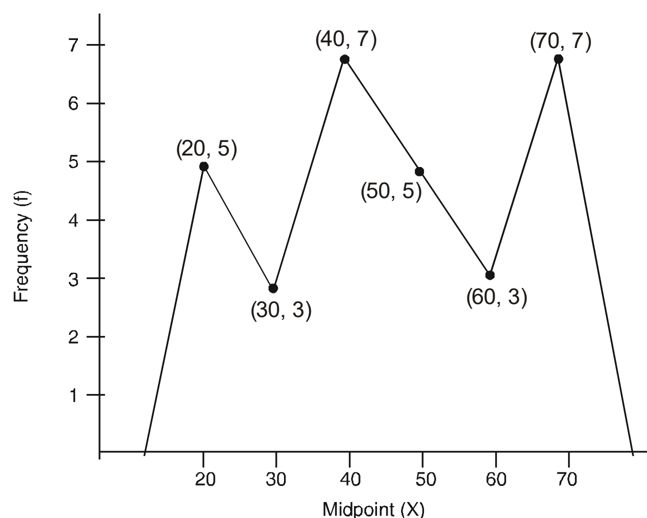
NOTES



Frequency Polygon

A frequency polygon is a line chart of frequency distribution in which either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines. Since the frequencies generally do not start at zero or end at zero, this diagram as such would not touch the horizontal axis. However, since the area under the entire curve is the same as that of a histogram which is 100 per cent of the data presented, the curve can be enclosed so that the starting point is joined with a fictitious preceding point whose value is zero. This ensures that the start of the curve is at horizontal axis and the last point is joined with a fictitious succeeding point whose value is also zero, so that the curve ends at the horizontal axis. This enclosed diagram is known as the frequency polygon.

We can construct the frequency polygon from the table presented for the ages of 30 workers as follows:



NOTES

Relative Frequency

In a frequency distribution, if the frequency in each class interval is converted into a proportion by dividing by the total frequency we get a series of proportions called *relative frequencies*. A distribution presented with relative frequencies rather than actual frequencies is called a *relative frequency* distribution. The sum of all relative frequencies in a distribution is 1.

Example 2: Calculate relative frequency from the given table.

Class Interval	*Frequency*
25—35	7
35—45	9
45—55	22
55—65	7
65—75	3
75—85	2

Solution: This example shows that the sum of all relative frequencies in a distribution is 1.

Class Interval	*Frequency*	*Relative Frequency*	*Explanation*
25—35	7	0.14	$\frac{7}{50} = 0.14$
35—45	9	0.18	$\frac{9}{50} = 0.18$
45—55	22	0.44	etc.
55—65	7	0.14	
65—75	3	0.06	
75—85	2	0.04	
	Total 50	1.00	

The concept of relative frequencies is useful in sampling theory. It can also be used to compare two frequency distributions with unequal total frequency with the same series of class intervals as in the following example.

Example 3: Compare the following frequency distribution.

Class Interval	f_1	f_2
10—20	5	12
20—30	10	24
30—40	6	30
40—50	3	19
50—60	1	15

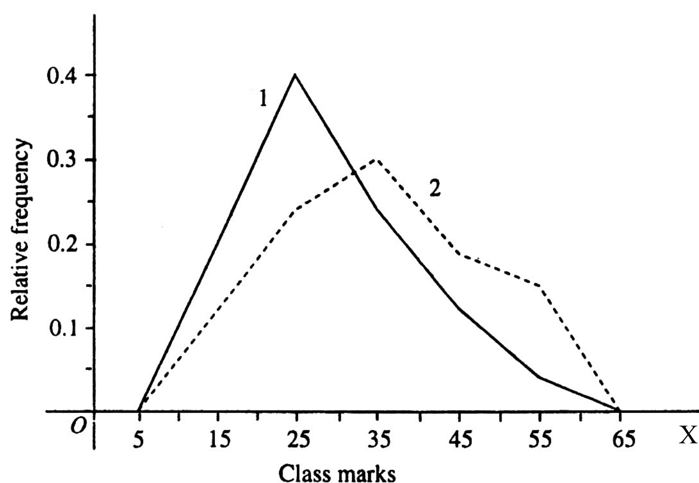
Solution: The following table shows the comparison.

Class Interval	f_1	f_2	*Rel. Freq.* f_1	*Rel. Freq.* f_2
10—20	5	12	0.20	0.12
20—30	10	24	0.40	0.24
30—40	6	30	0.24	0.30
40—50	3	19	0.12	0.19
50—60	1	15	0.04	0.15
Total	25	100	1.00	1.00

A direct visual comparison of two frequency distributions can be made by drawing their frequency polygons.

Example 4: Draw frequency polygons for the relative frequency distributions given in Example 3.

Solution: The following is the frequency polygon for the relative frequencies as mentioned in Example 3



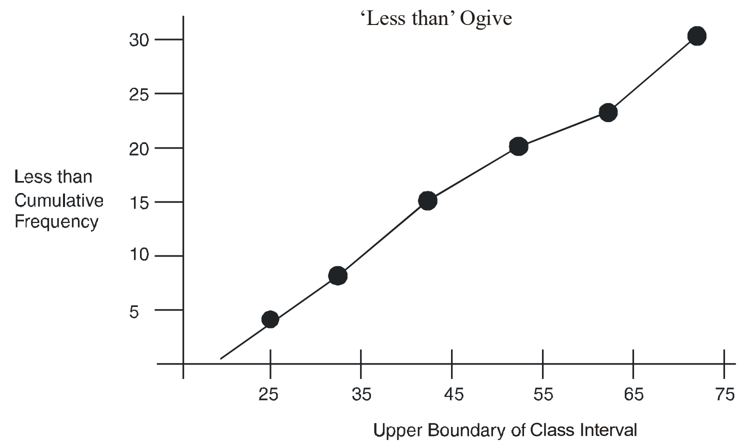
Ogive Curves

Cumulative frequency curve or ogive is the graphic representation of a cumulative frequency distribution. Ogives are of two types. One of these is less than and the other one is greater than ogive. Both these ogives are constructed based upon the following table of our example of 30 workers.

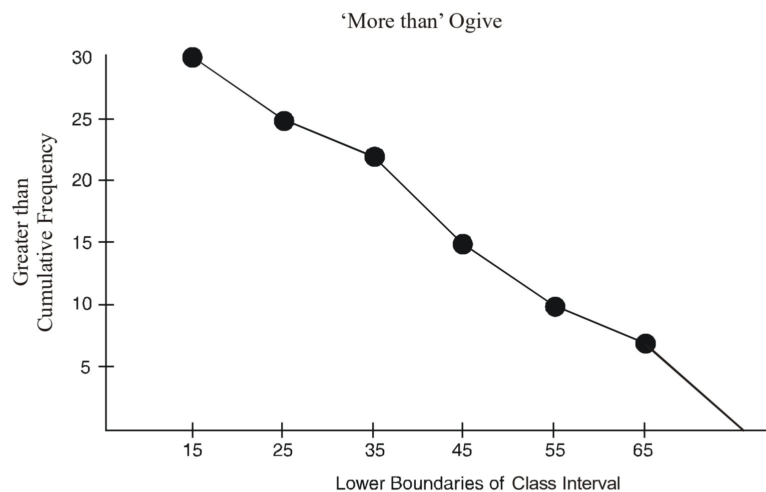
NOTES

Class Interval (years)	Mid-point	(f)	Cum. Freq. (less than)	Cum. Freq. (greater than)
15 and upto 25	20	5	5 (less than 25)	30 (more than 15)
25 and upto 35	30	3	8 (less than 35)	25 (more than 25)
35 and upto 45	40	7	15 (less than 45)	22 (more than 35)
45 and upto 55	50	5	20 (less than 55)	15 (more than 45)
55 and upto 65	60	3	23 (less than 65)	10 (more than 55)
65 and upto 75	70	7	30 (less than 75)	7 (more than 65)

(*i*) **Less than Ogive.** In this case, the less than cumulative frequencies are plotted against the upper boundaries of their respective class intervals.



(*ii*) **Greater than Ogive.** In this case, the greater than cumulative frequencies are plotted against the lower boundaries of their respective class intervals.



These ogives can be used for comparison purposes. Several ogives can be drawn on the same grid, preferably with different colours for easier visualization and differentiation.

Although, diagrams and graphs are powerful and effective media for presenting statistical data, they can only represent a limited amount of information and they are not of much help when intensive analysis of data is required.

Histograms

A histogram is the graphical description of data and is constructed from a frequency table. It displays the distribution method of a data set and is used for statistical as well as mathematical calculations.

The word histogram is derived from the Greek word histos which means 'anything set upright' and 'gramma' which means 'drawing, record, writing'. It is considered as the most important basic tool of statistical quality control process.

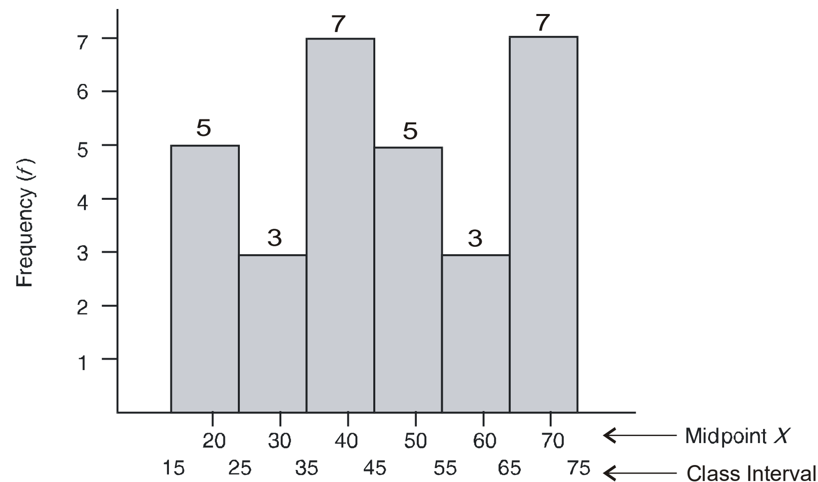
In this type of representation, the given data is plotted in the form of a series of rectangles. Class intervals are marked along the *X*-axis and the frequencies along the *Y*-axis according to a suitable scale. Unlike the bar chart, which is one-dimensional, meaning that only the length of the bar is important and not the width, a histogram is two-dimensional in which both the length and the width are important. A histogram is constructed from a frequency distribution of a grouped data, where the height of the rectangle is proportional to the respective frequency and the width represents the class interval. Each rectangle is joined with the other and any blank spaces between the rectangles would mean that the category is empty and there are no values in that class interval.

As an example, let us construct a histogram for our example of ages of 30 workers. For convenience sake, we will present the frequency distribution along with the midpoint of each interval, where the midpoint is simply the average of the values of the lower and the upper boundary of each class interval. The frequency distribution table is shown as follows:

Class Interval (years)	Midpoint	(f)
15 and upto 25	20	5
25 and upto 35	30	3
35 and upto 45	40	7
45 and upto 55	50	5
55 and upto 65	60	3
65 and upto 75	70	7

NOTES

The histogram of this data would be shown as follows:



Check Your Progress

1. What is a frequency polygon?
2. What are the two types of ogives?

9.3 CALCULATION OF MEAN

A weakness of the measures of dispersion, based upon the range or a portion thereof, is that the precise size of most of the variants has no effect on the result. As an illustration, the quartile deviation will be the same whether the variates between Q_1 and Q_3 are concentrated just above Q_1 or they are spread uniformly from Q_1 to Q_3. This is an important defect from the viewpoint of measuring the divergence of the distribution from its typical value. The mean deviation is employed to answer the objection.

Mean deviation, also called average deviation, of a frequency distribution is the mean of the absolute values of the deviation from some measure of central tendency. In other words, mean deviation is the arithmetic average of the variations (deviations) of the individual items of the series from a measure of their central tendency.

We can measure the deviations from any measure of central tendency, but the most commonly employed ones are the median and the mean. The median is preferred because it has the important property that the average deviation from it is the least.

Calculation of mean deviation then involves the following steps:

(*i*) Calculate the median (or the mean) *Me* (or $\overline{X}$).

(*ii*) Record the deviations $|d| = |x - Me|$ of each of the items, ignoring the sign.

(*iii*) Find the average value of deviations.

NOTES

$$\text{Mean Deviation} = \frac{\Sigma |d|}{N}$$

Example 5: Calculate the mean deviation from the following data giving marks obtained by 11 students in a class test.

14, 15, 23, 20, 10, 30, 19, 18, 16, 25, 12

Solution: Median = Size of $\frac{11+1}{2}$ th item

= Size of 6th item = 18

Serial No.	*Marks*	*| x – Median |* *| d |*
1	10	8
2	12	6
3	14	4
4	15	3
5	16	2
6	18	0
7	19	1
8	20	2
9	23	5
10	25	7
11	30	12
		$\sum \|d\| = 50$

$$\text{Mean Deviation from Median} = \frac{\sum |d|}{N}$$

$$= \frac{50}{11} = 4.5 \text{ marks}$$

For grouped data, it is easy to see that the mean deviation is given by,

$$\text{Mean Deviation (M.D.)} = \frac{\sum f|d|}{\sum f}$$

Here, $|d| = |x - \text{median}|$ for grouped discrete data, and $|d| = M - \text{median}|$ for grouped continuous data with M as the mid-value of a particular group. The following examples illustrate the use of this formula.

Example 6: Calculate the mean deviation from the following data:

Size of Item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

NOTES

Solution:

Size	*Frequency* (f)	*Cumulative frequency*	*Deviations from median* $\|d\|$	$f\|d\|$
6	3	3	3	9
7	6	9	2	12
8	9	18	1	9
9	13	31	0	0
10	8	39	1	8
11	5	44	2	10
12	4	48	3	12
	48			60

Median = Size of $\frac{48+1}{2}$ = 24.5th item which is 9

Therefore, deviations d are calculated from 9, i.e., $|d| = |x - 9|$.

Mean Deviation = $\frac{\sum f|d|}{\sum f} = \frac{60}{48} = 1.25$

Example 7: Calculate the mean deviation from the following data:

x	0–10	10–20	20–30	30–40	40–50	50–60	60–70	70–80
f	18	16	15	12	10	5	2	2

Solution:

This is a frequency distribution with continuous variable. Thus, deviations are calculated from mid-values.

x	*Mid-value*	f	*Less than c.f.*	*Deviation from median* $\|d\|$	$f\|d\|$
0–10	5	18	18	19	342
10–20	15	16	34	9	144
20–30	25	15	49	1	15
30–40	35	12	61	11	132
40–50	45	10	71	21	210
50–60	55	5	76	31	155
60–70	65	2	78	41	82
70–80	75	2	80	51	102
		80			1182

Median = Size of $\frac{80}{2}$ th item

$= 20 + \frac{6}{15} \times 10 = 24$

Hence, Mean Deviation $= \dfrac{\Sigma f |d|}{\Sigma f}$

$$= \frac{1182}{80} = 14.775$$

NOTES

Merits and Demerits of the Mean Deviation

Merits

1. It is easy to understand.
2. As compared to standard deviation (discussed later), its computation is simple.
3. As compared to standard deviation, it is less affected by extreme values.
4. Since it is based on all values in the distribution, it is better than range or quartile deviation.

Demerits

1. It lacks those algebraic properties which would facilitate its computation and establish its relation to other measures.
2. Due to this, it is not suitable for further mathematical processing.

Coefficient of Mean Deviation

The coefficient or relative dispersion is found by dividing the mean deviations recorded. Thus,

$$\text{Coefficient of M.D.} = \frac{\text{Mean Deviation (M.D.)}}{\text{Mean}}$$

(when deviations were recorded from the mean)

$$= \frac{\text{M.D.}}{\text{Median}}$$

(when deviations were recorded from the median)

Applying this formula to Example 5.3.

$$\text{Coefficient of Mean Deviation} = \frac{14.775}{24}$$

$$= 0.616$$

9.4 STANDARD DEVIATION

By far, the most universally used and the most useful measure of dispersion is the standard deviation or the root mean square deviation about the mean. We have seen that all the methods of measuring dispersion so far discussed are not universally adopted for want of adequacy and accuracy. The range is not satisfactory as its magnitude is determined by most extreme cases in the entire group. Further, the

NOTES

range is notable because it is dependent on the item whose size is largely a matter of chance. Mean deviation method is also an unsatisfactory measure of scatter, as it ignores the algebraic signs of deviation. We desire a measure of scatter which is free from these shortcomings. To some extent standard deviation is one such measure.

The calculation of standard deviation differs in the following respects from that of mean deviation. First, in calculating standard deviation, the deviations are squared. This is done so as to get rid of negative signs without committing algebraic violence. Further, the squaring of deviations provides added weight to the extreme items, a desirable feature for certain types of series.

Second, the standard deviations are always recorded from the arithmetic mean, because although the sum of deviations is the minimum from the median, the sum of squares of deviations is minimum when deviations are measured from the arithmetic average. The deviation from $\bar{x}$ is represented by d.

Thus, standard deviation, σ (sigma) is defined as the square root of the mean of the squares of the deviations of individual items from their arithmetic mean.

$$\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$$

For grouped data (discrete variables)

$$\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$$

and, for grouped data (continuous variables)

$$\sigma = \sqrt{\frac{\Sigma f(M-\bar{x})}{\Sigma f}}$$

Where, M is the mid-value of the group.

The use of these formulae is illustrated by the following examples.

Example 8: Compute the standard deviation for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21

Solution: Here the formula $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$ is appropriate. We first calculate the mean as $\bar{x} = \Sigma x/N = 176/11 = 16$, and then calculate the deviation as follows:

x	$(x-\bar{x})$	$(x-\bar{x})^2$
11	–5	25
12	–4	16
13	–3	9
14	–2	4
15	–1	1

NOTES

16	0	0
17	+1	1
18	+2	4
19	+3	9
20	+4	16
21	+5	25
176		11

Thus, by using the formula, $\sigma = \sqrt{\frac{\Sigma(x-\bar{x})^2}{N}}$, we get

$$\sigma = \sqrt{\frac{110}{11}} = \sqrt{10} = 3.16$$

Example 9: Find the standard deviation of the data in the following distributions:

x	12	13	14	15	16	17	18	20
f	4	11	32	21	15	8	6	4

Solution: For this discrete variable grouped data, we use the formula $\sigma = \sqrt{\frac{\Sigma f(x-\bar{x})^2}{\Sigma f}}$. Since for calculation of $\bar{x}$, we need Σfx and then for σ we need $\Sigma f(x-\bar{x})^2$, the calculations are conveniently made in the following format.

x	f	fx	$d = x - \bar{x}$	d^2	fd^2
12	4	48	–3	9	36
13	11	143	–2	4	44
14	32	448	–1	1	32
15	21	315	0	0	0
16	15	240	1	1	15
x	f	fx	$d = x - \bar{x}$	d^2	fd^2
17	8	136	2	4	32
18	5	90	3	9	45
20	4	80	5	25	100
	100	1500			304

Here, $\bar{x} = \Sigma fx/\Sigma f = 1500/100 = 15$

and $\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f}} = \sqrt{\frac{304}{100}} = \sqrt{3.04} = 1.74$

Example 10: Calculate the standard deviation of the following data:

Class	1–3	3–5	5–7	7–19	9–11	11–13	13–15
Frequency	1	9	25	35	17	10	3

Solution: This is an example of continuous frequency series and the formula $\sigma = \sqrt{\frac{\Sigma f(M-\bar{x})}{\Sigma f}}$ seems appropriate.

NOTES

Class	*Mid-point* (*x*)	*Frequency* (*f*)	*f* (*x*)	*Deviation of mid-point x from mean* (*d*)	*Squared deviation* d^2	*Squared deviation times frequency* fd^2
1–3	2	1	2	–6	36	36
3–5	4	9	36	–4	16	144
5–7	6	25	150	–2	4	100
7–9	8	35	280	0	0	0
9–11	10	17	170	2	4	68
11–13	12	10	120	4	16	160
13–15	14	3	42	6	36	108
		100	800			616

First the mean is calculated as,

$$\bar{x} = \Sigma fx/\Sigma f = 800/100 = 8.0$$

Then the deviations are obtained from 8.0. The standard deviation,

$$\sigma = \sqrt{\frac{\Sigma f(M-\bar{x})^2}{\Sigma f}}$$

$$\sigma = \sqrt{\frac{\Sigma fd^2}{\Sigma f}} = \sqrt{\frac{616}{100}}$$

$$= 2.48$$

Calculation of Standard Deviation by Short-Cut Method

The three examples worked out previously have one common simplifying feature, namely $\bar{x}$ in each, turned out to be an integer, thus simplifying calculations. In most cases, it is very unlikely that it will turn out to be so. In such cases, the calculation of d and d^2 becomes quite time-consuming. Short-cut methods have consequently been developed. These are on the same lines as those for the calculation of mean itself.

In the short-cut method, we calculate deviations x' from an assumed mean A. Then, for ungrouped data

$$\sigma = \sqrt{\frac{\Sigma x'^2}{N} - \left(\frac{\Sigma x'}{N}\right)^2}$$

and for grouped data

$$\sigma = \sqrt{\frac{\Sigma fx'^2}{\Sigma f} - \left(\frac{\Sigma fx'}{\Sigma f}\right)^2}$$

NOTES

This formula is valid for both discrete and continuous variables. In case of continuous variables, x in the equation $x' = x - A$ stands for the mid-value of the class in question.

Note that the second term in each of the formulae is a correction term because of the difference in the values of A and $\bar{x}$. When A is taken as $\bar{x}$ itself, this correction is automatically reduced to zero. The following examples explain the use of these formulae.

Example 11: Compute the standard deviation by the short-cut method for the following data:

11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21

Solution: Let us assume that $A = 15$.

	$x' = (x - 15)$	x'^2
11	–4	16
12	–3	9
13	–2	4
14	–1	1
15	0	0
16	1	1
17	2	4
18	3	9
19	4	16
20	5	25
21	6	36
$N = 11$	$\sum x' = 11$	$\sum x'^2 = 121$

$$\sigma = \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

$$= \sqrt{\frac{121}{11} - \left(\frac{11}{11}\right)^2}$$

$$= \sqrt{11 - 1}$$

$$= \sqrt{10}$$

$$= 3.16$$

Another Method: If we assumed A as zero, then the deviation of each item from the assumed mean is the same as the value of item itself. Thus, 11 deviates from the assumed mean of zero by 11, 12 deviates by 12, and so on. As such, we work with deviations without having to compute them, and the formula takes the following shape:

x	x^2
11	121
12	144
13	169
14	196
15	225

NOTES

16	256
17	289
18	324
19	361
20	400
21	441
176	2,926

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

$$= \sqrt{\frac{2926}{11} - \left(\frac{176}{11}\right)^2} = \sqrt{266 - 256} = 3.16$$

Example 12: Calculate the standard deviation of the following data by short-cut method.

Person	1	2	3	4	5	6	7
Monthly Income (Rupees)	300	400	420	440	460	480	580

Solution: In this data, the values of the variables are very large making calculations cumbersome. It is advantageous to take a common factor out. Thus, we use $x' = \frac{x - A}{20}$. The standard deviation is calculated using x' and then the true value of σ is obtained by multiplying back by 20. The effective formula used is,

$$\sigma = C \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2}$$

Here, C represents the common factor.

Using $x' = (x - 420)/20$

x	*Deviation from assumed mean* $x' = (x - 420)$	x'	x'^2
300	–120	–6	36
400	–20	–1	1
420	0	0	0
		–7	
440	20	1	1
460	40	2	4
480	60	3	9
580	160	8	64
		+14	
$N = 7$		$\Sigma x' = 7$	$\Sigma x'^2 = 115$

$$\sigma = 20 \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum x'}{N}\right)^2} = 20\sqrt{\frac{115}{7} - \left(\frac{7}{7}\right)^2} = 78.56$$

Example 13: Calculate the standard deviation from the following data:

Size	6	9	12	15	18
Frequency	7	12	19	10	2

Solution:

x	*Frequency* (f)	*Deviation from assumed mean 12*	*Deviation divided by common factor 3* (x')	x' *times frequency* (fx')	x'^2 *times frequency* (fx'^2)
6	7	–6	–2	–14	28
9	12	–3	–1	–12	12
12	19	0	0	0	0
15	10	3	1	10	10
18	2	6	2	4	8
	$N = 50$			$\sum fx' = -12$	$\sum fx'^2 = 58$

Since deviations have been divided by a common factor, we use,

$$\sigma = C\sqrt{\frac{\sum fx'^2}{N} - \left(\frac{\sum fx'}{N}\right)^2}$$

$$= 3\sqrt{\frac{58}{50} - \left(\frac{-12}{50}\right)^2}$$

$$= 3\sqrt{1.1600 - 0.0576}$$

$$= 3 \times 1.05 = 3.15$$

Example 14: Obtain the mean and the standard deviation of the first N natural numbers, i.e., of 1, 2, 3, ..., $N - 1$, N.

Solution: Let x denote the variable which assumes the values of the first N natural numbers.

Then,

$$\bar{x} = \frac{\sum_1^N x}{N} = \frac{\frac{N(N+1)}{2}}{N} = \frac{N+1}{2}$$

Hence, $$\sum_1^N x = 1 + 2 + 3 + ... + (N-1) + N$$

$$= \frac{N(N+1)}{2}$$

NOTES

To calculate the standard deviation σ, we use 0 as the assumed mean A. Then,

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2}$$

But $\quad \sum x^2 \quad = 1^2 + 2^2 + 3^2 + ... (N-1)^2 + N^2 = \dfrac{N(N+1)(2N+1)}{6}$

Therefore,

$$\sigma = \sqrt{\frac{N(N+1)(2N+1)}{6N} - \frac{N^2(N+1)^2}{4N^2}}$$

$$= \sqrt{\frac{(N+1)}{2}\left[\frac{2N+1}{3} - \frac{N+1}{2}\right]} = \sqrt{\frac{(N+1)(N-1)}{12}}$$

Thus, for the first 11 natural numbers

$$\bar{x} = \frac{11+1}{2} = 6$$

and $\quad \sigma = \sqrt{\dfrac{(11+1)(11-1)}{12}} = \sqrt{10} = 3.16$

Example 15:

	Mid-point (*x*)	*Frequency* (*f*)	*Deviation from class of assumed mean* (*x*′)	*Deviation time frequency* (*fx*′)	*Squared deviation times frequency* (*fx*′²)
0–10	5	18	–2	–36	72
10–20	15	16	–1	–16	16
				–52	
20–30	25	15	0	0	0
30–40	35	12	1	12	12
40–50	45	10	2	20	40
50–60	55	5	3	15	45
60–70	65	2	4	8	32
70–80	75	1	5	5	25
				60	
		$\Sigma f = 79$		60	242
				–52	
				$\sum fx' = 8$	

Solution: Since the deviations are from assumed mean and expressed in terms of class-interval units,

$$\sigma = i \times \sqrt{\frac{\sum x'^2}{N} - \left(\frac{\sum fx'}{N}\right)^2}$$

$$= 10 \times \sqrt{\frac{242}{79} - \left(\frac{8}{79}\right)^2}$$

$$= 10 \times 1.75 = 17.5$$

Combining Standard Deviations of Two Distributions

If we were given two sets of data of N_1 and N_2 items with means $\bar{x}_1$ and $\bar{x}_2$ and standard deviations σ_1 and σ_2 respectively, we can obtain the mean and the standard deviation $\bar{x}$ and σ of the combined distribution by the following formulae:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$$

and $$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}}$$

Example 16: The mean and the standard deviations of two distributions of 100 and 150 items are 50, 5 and 40, 6 respectively. Find the standard deviation of all taken together.

Solution: Combined mean,

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2} = \frac{100 \times 50 + 150 \times 40}{100 + 150} = 44$$

Combined standard deviation,

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}}$$

$$= \sqrt{\frac{100 \times (5)^2 + 150\,(6)^2 + 100\,(44 - 50)^2 + 150\,(44 - 40)^2}{100 + 150}}$$

$$= 7.46$$

Example 17: A distribution consists of three components with 200, 250, 300 items having mean 25, 10 and 15 and standard deviation 3, 4 and 5, respectively. Find the standard deviation of the combined distribution.

Solution: In the usual notations, we are given here:

$$N_1 = 200, N_2 = 250, N_3 = 300$$

$$\bar{x}_1 = 25, \bar{x}_2 = 10, \bar{x}_3 = 15$$

The formulae $\bar{x} = \dfrac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$ and $\sigma = \sqrt{\dfrac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}}$

can easily be extended for combination of three series as

NOTES

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + N_3\bar{x}_3}{N_1 + N_2 + N_3}$$

$$= \frac{200 \times 25 + 250 \times 10 + 300 \times 15}{200 + 250 + 300}$$

$$= \frac{12000}{750} = 16$$

and

$$\sigma = \sqrt{\frac{N_1\sigma_1^2 + N_2\sigma_2^2 + N_3\sigma_3^2 + N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2 + N_3(\bar{x} - \bar{x}_3)^2}{N_1 + N_2 + N_3}}$$

$$= \sqrt{\frac{200 \times 9 + 250 \times 16 + 300 \times 25 + 200 \times 81 + 250 \times 36 + 300 \times 1}{200 + 250 + 300}}$$

$$= \sqrt{51.73} = 7.19$$

Comparison of Various Measures of Dispersion

The range is the easiest to calculate the measure of dispersion, but since it depends on extreme values, it is extremely sensitive to the size of the sample, and to the sample variability. In fact, as the sample size increases the range increases dramatically, because the more the items one considers, the more likely it is that some item will turn up which is larger than the previous maximum or smaller than the previous minimum. So, it is, in general, impossible to interpret properly the significance of a given range unless the sample size is constant. It is for this reason that there appears to be only one valid application of the range, namely in statistical quality control where the same sample size is repeatedly used, so that comparison of ranges are not distorted by differences in sample size.

The quartile deviations and other such positional measures of dispersions are also easy to calculate, but suffer from the disadvantage that they are not amenable to algebraic treatment. Similarly, the mean deviation is not suitable because we cannot obtain the mean deviation of a combined series from the deviations of component series. However, it is easy to interpret and easier to calculate than the standard deviation.

The standard deviation of a set of data, on the other hand, is one of the most important statistics describing it. It lends itself to rigorous algebraic treatment, is rigidly defined and is based on all observations. It is, therefore, quite insensitive to sample size (provided the size is 'large enough') and is least affected by sampling variations.

NOTES

It is used extensively in testing of hypothesis about population parameters based on sampling statistics.

In fact, the standard deviation has such stable mathematical properties that it is used as a standard scale for measuring deviations from the mean. If we are told that the performance of an individual is 10 points better than the mean, it really does not tell us enough, for 10 points may or may not be a large enough difference to be of significance. But if we know that the σ for the score is only 4 points, so that on this scale, the performance is 2.5 σ better than the mean, the statement becomes meaningful. This indicates an extremely good performance. This sigma scale is a very commonly used scale for measuring and specifying deviations which immediately suggest the significance of the deviation.

The only disadvantage of the standard deviation lies in the amount of work involved in its calculation, and the large weight it attaches to extreme values because of the process of squaring involved in its calculations.

Check Your Progress

3. Explain mean deviation.
4. What are the merits of mean deviation.
5. Define standard deviation.
6. Calculate the standard deviation for the series 1, 2, 3, 5, 7.
7. For a group of 50 male workers, the mean and standard deviation of their weekly wages are Rs 63 and Rs 9 respectively. For a group of 40 female workers these are Rs 54 and 6 respectively. Find the standard deviation of the combined group of 90 workers.
8. (a) Mean and standard deviations of two distributions of 100 and 150 item are 50, 5 and 40, 6 respectively. Find the mean and standard deviations of all the 250 items taken together.

   (b) Mean and standard deviations of 100 items are found by a student as 9 and 5. If at the time of calculations two items are wrongly taken as 40 and 50 instead of 60 and 30, find the correct mean and standard deviations.

9.5 COEFFICIENT OF VARIATION

The square of standard deviation, namely σ^2, is termed as variance and is more often specified than the standard deviation. Clearly, it has the same properties as standard deviation.

As is clear, the standard deviation σ or its square, the variance, cannot be very useful in comparing two series where either the units are different or the mean values are different. Thus, a σ of 5 on an examination where the mean score is 30 has an altogether different meaning than on an examination where the mean score

is 90. Clearly, the variability in the second examination is much less. To take care of this problem, we define and use a coefficient of variation, V,

NOTES

$$V = \frac{\sigma}{\bar{x}} \times 100$$

expressed as percentage.

Example 18: The following are the scores of two batsmen A and B in a series of innings:

A	12	115	6	73	7	19	119	36	84	29
B	47	12	76	42	4	51	37	48	13	0

Who is the better run-getter? Who is more consistent?

Solution: In order to decide as to which of the two batsmen, A and B, is the better run-getter, we should find their batting averages. The one whose average is higher will be considered as a better batsman.

To determine the consistency in batting we should determine the coefficient of variation. The less this coefficient the more consistent will be the player.

A			*B*		
Scores *x*	*x*	x^2	*Scores* *x*	*x*	x^2
12	–38	1,444	47	14	196
115	+65	4,225	12	–21	441
6	–44	1,936	76	43	1,849
73	+23	529	42	9	81
7	–43	1,849	–4	– 29	841
19	–31	961	51	18	324
119	+69	4,761	37	4	16
36	–14	196	48	15	225
84	+34	1,156	13	–20	400
29	–21	441	0	–33	1,089
$\sum x = 500$		17,498	$\sum x = 330$		5,462

Batsman A:

$$\bar{x} = \frac{500}{10} = 50$$

$$\sigma = \sqrt{\frac{17,498}{10}} = 41.83$$

$$V = \frac{41.83 \times 100}{50}$$

$$= 83.66 \text{ per cent}$$

Batsman B:

$$\bar{x} = \frac{330}{10} = 33$$

$$\sigma = \sqrt{\frac{5,462}{10}} = 23.37$$

$$V = \frac{23.37}{33} \times 100$$

$$= 70.8 \text{ per cent}$$

A is a better batsman since his average is 50 as compared to 33 of B. But B is more consistent since the variation in his case is 70.8 as compared to 83.66 of A.

Example 19: The following table gives the age distribution of students admitted to a college in the years 1914 and 1918. Find which of the two groups is more variable in age.

NOTES

Age	*Number of Students in*	
	1914	*1918*
15	–	1
16	1	6
17	3	34
18	8	22
19	12	35
20	14	20
21	13	7
22	5	19
23	2	3
24	3	–
25	1	–
26	–	–
27	1	–

Solution:

Age	*Assumed Mean–21 1914*				*Assumed Mean–19 1918*			
	f	*x′*	*fx′*	*fx′²*	*f*	*x′*	*fx*	*fx′²*
15	0	–6	0	0	1	–4	–4	16
16	1	–5	–5	25	6	–3	–18	54
17	3	–4	–12	48	34	–2	–68	136
18	8	–3	–24	72	22	–1	–22	22
19	12	–2	–24	48			–112	
20	14	–1	–14	14				
			–79		35	0	0	0
21	13	0	0	0	20	1	20	20
22	5	1	5	5	7	2	14	28
23	2	2	4	8	19	3	57	171
24	3	3	9	27	3	4	12	48
25	1	4	4	16	147		+103	495
26	0	5	0	0			–9	
27	1	6	6	36				
	63		+28	299				
			–51					

NOTES

1914 Group:

$$\sigma = \sqrt{\frac{\sum fx'^2}{N} - \left[\frac{\sum (fx')}{N}\right]^2}$$

$$= \sqrt{\frac{299}{63} - \left(\frac{-51}{63}\right)^2}$$

$$= \sqrt{4.476 - 0.655} = \sqrt{4.091}$$

$$= 2.02$$

$$\bar{x} = 21 + \left(\frac{-51}{63}\right) = 21 - 8 = 20.2$$

$$V = \frac{2.02}{20.2} \times 100$$

$$= \frac{202}{20.2} = 10$$

1918 Group:

$$\sigma = \sqrt{\frac{495}{147} - \left(\frac{-9}{147}\right)^2} = \sqrt{3.3673 - 0.0037}$$

$$= \sqrt{3.3636} = 1.834$$

$$\bar{x} = 19 + \left(\frac{-9}{147}\right)$$

$$= 19 - .06 = 18.94$$

$$V = \frac{1.834}{18.94} \times 100$$

$$= 9.68$$

The coefficient of variation of the 1914 group is 10 and that of the 1918 group is 9.68. This means that the 1914 group is more variable, but only barely so.

Example 20: You are supplied the following data about the height of boys and girls studying in a college.

	Boys	Girls
Number	72	38
Average height (inches)	68	61
Variance of distribution	9	4

You are required to find out:

(*i*) In which sex, boys or girls, is there greater variability in individual heights.

(*ii*) Common average height in boys and girls.

(*iii*) Standard deviation of the height of boys and girls taken together.

(*iv*) Combined variability (C.V.).

Solution:

(*i*) C.V. of boys' height $= \frac{\sigma_1}{\bar{x}_1} \times 100 = \frac{\sqrt{9}}{68} \times 100 = 4.41\%$

C.V. of girls' height $= \frac{\sigma_2}{\bar{x}_2} \times 100 = \frac{\sqrt{4}}{61} \times 100 = 3.28\%$

Thus, there is a greater variability in the height of boys than that of the girls.

(*ii*) Combined height of boys and girls is given as,

$$\bar{x}_{12} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N_1 + N_2}$$

$$= \frac{72 \times 68 + 38 \times 61}{72 + 38} = \frac{7214}{110} = 65.58 \text{ inches approx.}$$

(*iii*) The combined standard deviation may be calculated by applying the following formula:

$$\sigma_{12}^2 = \frac{N_1\sigma_1^2 + N_2\sigma_2^2}{N_1 + N_2} + \frac{N_1(\bar{x} - \bar{x}_1)^2 + N_2(\bar{x} - \bar{x}_2)^2}{N_1 + N_2}$$

$$= \frac{72 \times 9 + 38 \times 4}{72 + 38} + \frac{72(65.58 - 68)^2 + 38(65.58 - 61)^2}{72 + 38}$$

$$= \frac{2018.794}{110} = 18.35$$

$$\sigma_{12} = 4.28 \text{ inches}$$

(*iv*) Combined variability $= \frac{\sigma}{\bar{x}} \times 100 = \frac{4.28}{65.58} \times 100 = 6.53$

NOTES

Check Your Progress

9. Coefficients of variation of two series are 58% and 69%. Their standard deviations are 21.2 and 15.6. What are their arithmetic mean?
10. When can coefficient of variations be greater than 100%? What can you say about the items of the given data in such a case?

9.6 CORRELATION COEFFICIENT

Correlation analysis is the statistical tool generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomena which have relation to each other. For instance, when demand of a certain commodity increases, then its price goes up and when its demand decreases then its price comes down. Similarly, with age the height of the children increases; with height the weight of the children increases, with money supply the general level of prices go up. Such sort

NOTES

of relationship can as well be noticed for several other phenomena. The theory by means of which quantitative connections between two sets of phenomena are determined is called the *Theory of Correlation.*

On the basis of the theory of correlation one can study the comparative changes occurring in two related phenomena and their cause-effect relation can be examined. It should, however, be borne in mind that relationship like 'black cat causes bad luck', 'filled-up pitchers result in good fortune' and similar other beliefs of the people cannot be explained by the theory of correlation since they are all imaginary and are incapable of being justified mathematically. Thus, correlation is concerned with the relationship between two related and quantifiable variables. If two quantities vary in sympathy so that a movement (an increase or decrease) in the one tends to be accompanied by a movement in the same or opposite direction in the other and the greater the change in the one, the greater is the change in the other, the quantities are said to be correlated. This type of relationship is known as correlation or what is sometimes called, in statistics, co-variation.

For correlation it is essential that the two phenomena, should have cause-effect relationship. If such relationship does not exist then one should not talk of correlation. For example, if the height of the students as well as the height of the trees increases, then one should not call it a case of correlation because the two phenomena, viz., the height of students and the height of trees are not causally related. But the relationship between the price of a commodity and its demand, the price of a commodity and its supply, the rate of interest and savings, etc., are examples of correlation since in all such cases the change in one phenomenon is explained by a change in the other phenomenon.

Types of Correlation

It is appropriate here to mention that correlation in case of phenomena pertaining to natural sciences can be reduced to absolute mathematical terms, e.g., heat always increases with light. But in phenomena pertaining to social sciences, it is often difficult to establish any absolute relationship between two phenomena. Hence, in social sciences we must take the fact of correlation being established if in a large number of cases, two variables always tend to move in the same or the opposite direction.

Correlation can either be positive or it can be negative. Whether correlation is positive or negative would depend upon the direction in which the variables are moving. If both variables are changing in the same direction, then correlation is said to be positive but when the variations in the two variables take place in opposite direction, the correlation is termed as negative. This can be explained as follows:

Changes in Independent Variable	*Changes in Dependent Variable*	*Nature of Correlation*
Increase (+)↑	Increase (+)↑	Positive (+)
Decrease (–)↓	Decrease (–)↓	Positive (+)
Increase (+)↑	Decrease (–)↓	Negative (–)
Decrease (–)↓	Increase (+)↑	Negative (–)

NOTES

Correlation can either be linear or it can be non-linear. Non-linear correlation is also known as curvilinear correlation. The distinction is based upon the constancy of the ratio of change between the variables. When the amount of change in one variable tends to bear a constant ratio to the amount of change in the other variable then the correlation is said to be linear. In such a case, if the values of the variables are plotted on a graph paper, then a straight line is obtained. This is why the correlation is known as linear correlation. But when the amount of change in one variable does not bear a constant ratio to the amount of change in the other variable, i.e., the ratio happens to be variable instead of constant, then the correlation is said to be non-linear or curvilinear. In such a situation we shall obtain a curve if the values of the variables are plotted on a graph paper.

Correlation can either be simple correlation or it can be partial correlation or it can be multiple correlation. The study of correlation for two variables (of which one is independent and the other is dependent) involves application of simple correlation. When more than two variables are involved in a study relating to correlation then it can either be as of multiple correlation or of partial correlation. Multiple correlation studies the relationship between a dependent variable and two or more independent variables. In partial correlation, we measure the correlation between a dependent variable and one particular independent variable assuming that all other independent variables remain constant.

Statisticians have developed two measures for describing the correlation between two variables viz., the coefficient of determination and the coefficient of correlation.

Properties of Correlation Coefficient

The coefficient of correlation symbolically denoted by '*r*' is an important measure to describe how well one variable is explained by another. It measures the degree of relationship between the two causally-related variables. The value of this coefficient can never be more than + 1 or less than –1. Thus, + 1 and –1 are the limits of this coefficient. For a unit change in independent variable, if there happens to be a constant change in the dependent variable in the same direction then the value of the coefficient will be + 1 indicative of the perfect positive correlation; but if such a change occurs in the opposite direction, the value of the coefficient will be –1, indicating perfect negative correlation. In practical life the possibility of obtaining either a perfect positive or perfect negative correlation is very remote, particularly in respect of phenomena concerning social sciences. If the coefficient of correlation has a zero value then it means that there exists no correlation between the variables under study.

There are several methods of finding the coefficient of correlation but the following ones are considered important:

(*i*) Coefficient of correlation by the method of least squares

(*ii*) Coefficient of correlation through product moment method or Karl Pearson's coefficient of correlation

(*iii*) Coefficient of correlation using simple regression coefficients

Whichever of these three methods, we adopt we get the same value of *r*. Now, we explain in brief each one of these three methods of finding *'r'*.

9.6.1 Methods of Least Squares and Meaing Averages

NOTES

Least squares method of fitting a line (the line of best fit or the regression line) through the scatter diagram is a method which minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a fashion that the sum of the squares of the vertical deviations of these points from the line will be a minimum.

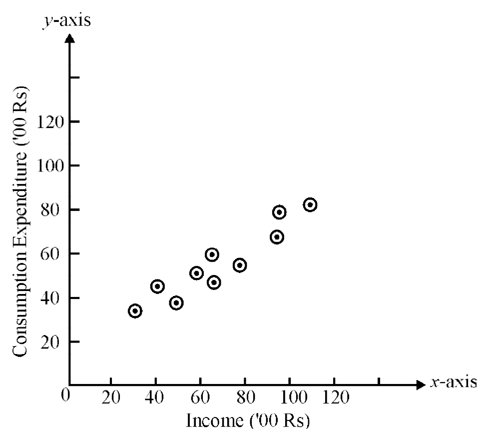


Fig 6.1 *Scatter Diagram*

The meaning of the least squares criterion can be better understood more easily through reference to the following figure where the Scatter diagram[2] has been reproduced along with a line which represents the least squares fit to the data.

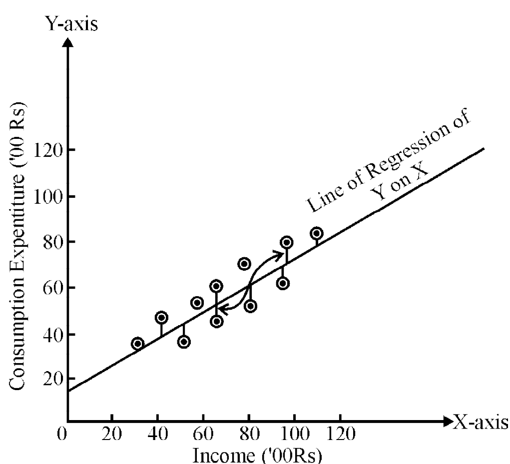


Fig 6.2 *Scatter Diagram, Regression Line and Short Vertical Lines Representing 'e'*

In the Figure 9.2, the vertical deviations of the individual points from the line are shown as the short vertical lines joining the points to the least squares line. These deviations are denoted by the symbol '*e*'. The value '*e*' varies from one

point to another. In some cases it is positive, in others it is negative. If the line drawn happens to be the least squares line then the values of $\sum e_i^o$ is the least possible. It is because of this feature the method is known as Least Squares Method.

Why we insist on minimizing the sum of squared deviations is a question that needs explanation. If we denote the deviations from the actual value Y to the estimated value as $(Y - \hat{Y})$ or e_i, it is logical that we want the $\Sigma(Y - \hat{Y})$ or $\sum_{i=1}^{n} e,$ to be as small as possible. However, mere examining $\Sigma(Y - \hat{Y})$ or $\sum_{i=1}^{n} e,$ is inappropriate since any e_i can be positive or negative and large positive values and large negative values would cancel one another.

But large values of e_i regardless of their sign, indicate a poor prediction. Even if we ignore the signs while working out $\sum_{i=1}^{n} | e_i |$, the difficulties may continue to be there. Hence, the standard procedure is to eliminate the effect of signs by squaring each observation. Squaring each term accomplishes two purposes viz., (*i*) It magnifies (or penalizes) the larger errors, and (*ii*) It cancels the effect of the positive and negative values (since a negative error squared becomes positive). The choice of minimizing the squared sum of errors rather than the sum of the absolute values implies that we would make many small errors rather than a few large errors. Hence, in obtaining the regression line we follow the approach that the sum of the squared deviations be minimum and on this basis work out the values of its constants viz., '*a*' and '*b*' or what is known as the intercept and the slope of the line. This is done with the help of the following two normal equations:[3]

$$\Sigma Y = na + b\Sigma X$$

$$\Sigma XY = a\Sigma X + \mathrm{b}\Sigma X^2$$

In these two equations, '*a*' and '*b*' are unknowns and all other values viz., $\sum X, \sum Y, \sum X^2, \sum XY$ are the sum of the products and the cross products to be calculated from the sample data and '*n*' means the number of observations in the sample. Hence, one can solve these two equations for finding the unknown values. Once these values are found, the regression line is said to have been defined for the given problem. Statisticians have also derived a short cut method through which these two equations can be rewritten so that the values of '*a*' and '*b*' can be directly obtained as follows:

$$b = \frac{n\sum XY - \sum X.\sum Y}{n\sum X^2 - \left(\sum X\right)^2}$$

$$a = \frac{\sum Y}{n} - b\frac{\Sigma X}{n}$$

NOTES

9.7 REGRESSION COEFFICIENT

Once a reasonable degree of correlation is established between two variables, we may show interest in estimating or predicting the value of one variable given the value of another. It is here that regression analysis comes into picture. Regression analysis reveals the average relationship between two variables and this makes possible estimation or prediction via a mathematical equation connecting the two variables.

1. Regression equation of X on Y:

$$(X - \bar{X}) = \frac{r.\sigma_x}{\sigma_y}(Y - \bar{Y})$$

$$\frac{r\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2}$$ (If deviations are taken from actual means of X and Y)

i.e., where $x = (X - \bar{X})$ and $y = (Y - \bar{Y})$

$$\frac{r\sigma_x}{\sigma_y} = \left[\frac{\Sigma d_x d_y - \frac{(\Sigma d_x)(\Sigma d_y)}{N}}{\Sigma d_y^2 - \frac{(\Sigma d_y)^2}{N}}\right]$$

(If deviations are taken from assumed means of X and Y)
i.e., if $d_x = (X - A_x)$ and $d_y = (y - A_y)$

2. Regression equation of Y on X:

$$(Y - \bar{Y}) = \frac{r.\sigma_y}{\sigma_x}(X - \bar{X})$$

$$\frac{r.\sigma_y}{\sigma_x} = \frac{\Sigma xy}{\Sigma x^2}$$

(If the deviations are taken from actual means of X and Y)

i.e., if $x = (X - \bar{X})$ and $y = (Y - \bar{Y})$

$$\frac{r.\sigma_y}{\sigma_x} = \left[\frac{\Sigma d_x d_y - \frac{(\Sigma d_x)(\Sigma d_y)}{N}}{\Sigma d_x^2 - \frac{(\Sigma d_x)^2}{N}}\right]$$

(If deviations are taken from assumed means of X and Y)
i.e., if $d_x = (X - A_x)$ and $d_y = (Y - A_y)$

3. Regression coefficients :

$\frac{r.\sigma_x}{\sigma_y}$ or b_{xy} is the regression coefficient of X on Y.

$\frac{r.\sigma_y}{\sigma_x}$ or b_{yx} is the regression coefficient of Y on X.

$$r = \sqrt{b_{xy} \times b_{yx}}$$

4. $\frac{r.\sigma_x}{\sigma_y} = \left(\frac{\Sigma xy}{N\sigma_x . \sigma_y}\right) \times \frac{\sigma_x}{\sigma_y} = \frac{\Sigma xy}{\Sigma y^2} = b_{xy}$

or $b_{xy} = \frac{r.\sigma_x}{\sigma_y} = \frac{\Sigma xy}{N.\sigma_y^2} = \frac{\mu_{11}}{\sigma_y^2}$ [where μ_{11} = Covariance (x, y)]

Note: In case we deal with actual values of X and Y variables and not the deviations,

then $b_{xy} = \left[\frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{N\Sigma Y^2 - (\Sigma Y)^2}\right]$

5. $b_{yx} = \frac{r.\sigma_y}{\sigma_x} = \left(\frac{\Sigma xy}{N\sigma_x . \sigma_y}\right) = \left(\frac{\sigma_y}{\sigma_x}\right)$

or $b_{yx} = \frac{\Sigma xy}{N\sigma_x^2} = \frac{\mu_{11}}{\sigma_x^2}$ [where μ_{11} = Covariance (x, y)]

Note: In case we deal with actual values of x and y variables and not the deviations,

then, $b_{yx} = \left[\frac{N(\Sigma XY) - (\Sigma X)(\Sigma Y)}{N\Sigma X^2 - (\Sigma X)^2}\right]$.

6. Angle between the lines of regression:
If θ be the acute angle between the two lines of regression (X on Y and Y on X, then,

$$\tan\theta = \left(\frac{1 - r^2}{r}\right) . \frac{\sigma_x \sigma_y}{\left(\sigma_x^2 + \sigma_y^2\right)}$$

Note:

(*i*) Both the lines of regression (of Y on X and X on Y) pass through the point $(\bar{X}, \bar{Y})$, the mean of X and Y series.

(*ii*) If the two lines of regression coincide, then the correlation between X and Y is perfect and by equating the respective slopes, we get,

$$\frac{r.\sigma_y}{\sigma_x} = \frac{\sigma_y}{r\sigma_x}$$

or $r^2 = 1$ or $r = \pm 1$

Hence, $\tan\theta = 0 \Rightarrow r^2 = 1 \Rightarrow r = \pm 1$

(*iii*) If the coefficient of correlation, viz., r between X and Y is zero, i.e., the variables X and Y are independent, it can be easily seen that the lines of regression of Y and X and of X on Y are respectively given by $Y = \bar{Y}$ and $X = \bar{X}$ and these two regression lines intersect at right angles.
Therefore, if $r = 0$, $\tan\theta = \infty \Rightarrow \theta = \pi/2$ or 90°.

NOTES

7. Standard error of estimate :
The standard error of regression of Y values from $Y_c = S_{yx}$

$$S_{yx} = \sqrt{\frac{\Sigma(Y - Y_c)^2}{N}} = \sqrt{\frac{\text{Unexplained variation}}{\text{N}}}$$

Also, $$S_{yx} = \sigma_y . \sqrt{(1-r^2)}$$

and $$S_{yx} = \frac{\sqrt{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}}{N}$$

Similarly, if S_{xy} stands for the standard error of regression of X values from X_c

then $$S_{xy} = \sqrt{\frac{\Sigma(X - X_c)^2}{N}}$$

$$S_{xy} = \sigma_x . \sqrt{(1-r^2)}$$

Also, $$S_{xy} = \sqrt{\frac{\Sigma X^2 - a\Sigma X - b\Sigma XY}{N}}$$

Note: The standard error of estimate measures the accuracy of the estimated figures. The smaller the value of standard error of estimate, the closer will be the dots to the regression line and the better the estimates based on the equation for this line. If the standard error of estimate is zero, then there is no variation about the line and the correlation will be perfect. Thus, with the help of standard error or estimate, it is possible for us to ascertain how good and representative the regression line is as a description of the average relationship between two series.

Properties of Regression Coefficient

(*i*) Coefficient of correlation r between the variables x and y is the geometric mean between two regression coefficients, b_{yx} and b_{xy}. (i.e., $r = \sqrt{b_{yx} \times b_{xy}}$).

(*ii*) Though $r_{xy} = r_{yx}$ (always), $b_{xy} \neq b_{yx}$ in general. (They become equal only when $\sigma_x^2 = \sigma_y^2$.)

(*iii*) If one of the regression coefficients is numerically, greater than unity then the other is less than unity.

(*iv*) The arithmetic mean of regression coefficients is greater than the coefficient of correlation r (by and large).

(*v*) The covariance, the coefficient of correlation r and the two regression coefficients have the same sign.

(*vi*) Though correlation coefficient is independent of both scale and origin, the regression coefficients are independent of change of origin but not of scale.

Check Your Progress

11. What are the different types of correlations?
12. Explain the meaning of correlation analysis.
13. What is the scatter diagram method?
14. What is the least-squares method?
15. List any two properties of regression coefficient.

NOTES

9.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. A frequency polygon is a line chart of frequency distribution in which either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines.
2. Ogives are of two types. One of these is less than and the other one is greater than ogive.
3. Mean deviation also called average deviation, of a frequency distribution is the mean of the absolute values of the deviation from some measure of central tendency. In other words, mean deviation is the arithmetic average of the variations (deviations) of the individual items of the series from a measure of their central tendency.
4. (*i*) It is easy to understand.

   (*ii*) As compared to standard deviation (discussed later), its computation is simple.

   (*iii*) As compared to standard deviation, it is less affected by extreme values.

   (*iv*) Since it is based on all values in the distribution, it is better than range or quartile deviation.
5. Standard deviation, σ (sigma) is defined as the square root of the mean of the squares of the deviations of individual items from their arithmetic mean.
6. 2.14
7. 4.09
8. (a) 7.46 (b) 5.4 approx
9. 36.6 approx., 22.6 approx.
10. Coefficient of variation A = 6.0% and coefficient of variation B = 6.4%. So Brick layer A will continue to be more consistent than brick layer B.
11. There are several types of correlations. They are:
    (a) Positive or negative correlations
    (b) Linear or non-linear correlations
    (c) Simple, partial or multiple correlations

NOTES

12. Correlation analysis is the statistical tool that is generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable. In fact, the word correlation refers to the relationship or interdependence between two variables. There are various phenomena, which are related to each other. For instance, when demand of a certain commodity increases, then its price goes up and when its demand decreases then its price comes down.
13. Scatter diagram is the method to calculate the constants in regression models that makes use of scatter diagram or dot diagram. A scatter diagram is a diagram that represents two series with the known variables, i.e., independent variable plotted on the *X*-axis and the variable to be estimated, i.e., dependent variable to be plotted on the *Y*-axis.
14. The least squares method is a method to calculate the constants in regression models for fitting a line through the scatter diagram that minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a fashion that the sum of the squares of the vertical deviations of these points from the line will be a minimum.
15. (*i*) If one of the regression coefficients is numerically greater than unity, then the other is less than unity.
    (*ii*) The arithmetic mean of regression coefficients is greater than the coefficient of correlation r.

9.9 SUMMARY

- In a graph, the independent variable should always be placed on the horizontal or *x*-axis and the dependent variable on the vertical or *y*-axis.
- A frequency polygon is a line chart of frequency distribution in which either the values of discrete variables or midpoints of class intervals are plotted against the frequencies and these plotted points are joined together by straight lines. Since the frequencies generally do not start at zero or end at zero, this diagram as such would not touch the horizontal axis.
- In a frequency distribution, if the frequency in each class interval is converted into a proportion by dividing by the total frequency we get a series of proportions called *relative frequencies*. A distribution presented with relative frequencies rather than actual frequencies is called a *relative frequency* distribution. The sum of all relative frequencies in a distribution is 1.
- Cumulative frequency curve or ogive is the graphic representation of a cumulative frequency distribution. Ogives are of two types. One of these is less than and the other one is greater than ogive.

- A histogram is the graphical description of data and is constructed from a frequency table. It displays the distribution method of a data set and is used for statistical as well as mathematical calculations.
- The word histogram is derived from the Greek word histos which means 'anything set upright' and 'gramma' which means 'drawing, record, writing'. It is considered as the most important basic tool of statistical quality control process.
- A weakness of the measures of dispersion, based upon the range or a portion thereof, is that the precise size of most of the variants has no effect on the result.
- Mean deviation, also called average deviation, of a frequency distribution is the mean of the absolute values of the deviation from some measure of central tendency. In other words, mean deviation is the arithmetic average of the variations (deviations) of the individual items of the series from a measure of their central tendency.
- By far, the most universally used and the most useful measure of dispersion is the standard deviation or the root mean square deviation about the mean.
- The standard deviations are always recorded from the arithmetic mean, because although the sum of deviations is the minimum from the median, the sum of squares of deviations is minimum when deviations are measured from the arithmetic average. The deviation from $\bar{x}$ is represented by d.
- The range is the easiest to calculate the measure of dispersion, but since it depends on extreme values, it is extremely sensitive to the size of the sample, and to the sample variability.
- The square of standard deviation, namely s^2, is termed as variance and is more often specified than the standard deviation. Clearly, it has the same properties as standard deviation.
- Correlation analysis is the statistical tool generally used to describe the degree to which one variable is related to another. The relationship, if any, is usually assumed to be a linear one. This analysis is used quite frequently in conjunction with regression analysis to measure how well the regression line explains the variations of the dependent variable.
- The theory by means of which quantitative connections between two sets of phenomena are determined is called the *Theory of Correlation.*
- Whether correlation is positive or negative would depend upon the direction in which the variables are moving. If both variables are changing in the same direction, then correlation is said to be positive but when the variations in the two variables take place in opposite direction, the correlation is termed as negative.
- The coefficient of correlation symbolically denoted by 'r' is an important measure to describe how well one variable is explained by another. It

NOTES

measures the degree of relationship between the two causally-related variables. The value of this coefficient can never be more than + 1 or less than –1. Thus, + 1 and –1 are the limits of this coefficient.

- Least squares method of fitting a line (the line of best fit or the regression line) through the scatter diagram is a method which minimizes the sum of the squared vertical deviations from the fitted line. In other words, the line to be fitted will pass through the points of the scatter diagram in such a fashion that the sum of the squares of the vertical deviations of these points from the line will be a minimum.
- Once a reasonable degree of correlation is established between two variables, we may show interest in estimating or predicting the value of one variable given the value of another. It is here that regression analysis comes into picture. Regression analysis reveals the average relationship between two variables and this makes possible estimation or prediction via a mathematical equation connecting the two variables.
- If one of the regression coefficients is numerically, greater than unity then the other is less than unity.
- The covariance, the coefficient of correlation r and the two regression coefficients have the same sign.

9.10 KEY WORDS

- **Mean deviation:** It is the arithmetic mean of the absolute deviations of a series of values.
- **Standard deviation:** It is the measure of the dispersion of a set of values and is calculated from the mean of squared deviations.
- **Correlation analysis:** It is the statistical tool to describe the degree to which one variable is related to another.

9.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. How will you present data in the graphic form?
2. What is a histogram?
3. List the different types of Ogives.
4. What are the ways in which data can be diagrammatically represented?
5. How will you calculate the mean deviation of a given data?

6. What is the significance of using standard deviation in statistical evaluation of data?
7. Define variance.
8. What is a scatter diagram?
9. Under what conditions will you say that a correlation is linear?
10. How does a scatter diagram help in studying the correlation between two variables?
11. List the different types of correlation.
12. Define correlation analysis.
13. Define regression analysis.

Long-Answer Questions

1. Calculate mean deviation and its coefficient about median, arithmetic mean and mode for the following figures, and show that the mean deviation about the median is least.

   103, 50, 68, 110, 108, 105, 174, 103, 150, 200, 225, 350, 103

2. Compute mean deviations of the two series and point out which is more variable.

Month	*Index No. Calcutta*	*Index No. Delhi*	*Month*	*Index No. Calcutta*	*Index No. Delhi*
1970 April	93	107	1970 October	97	107
1970 May	97	108	1970 November	97	105
1970 June	95	102	1970 December	92	101
1970 July	95	102	1971 January	93	100
1970 August	95	102	1971 February	89	97
1970 September	95	104	1971 March	89	96

3. Calculate (a) Median coefficient of dispersion and (b) Mean coefficient of dispersion from the following data:

Size of Items	14	16	18	20	22	24	26
Frequency	2	4	5	3	2	1	4

4. Compute the mean deviation from the median and from the mean for the following distribution of the scores of 50 college students. Also complete the class interval.

Scores	140–	150–	160–	170–	180–	190–200
Frequency	4	6	10	18	9	3

5. Find the mean deviation about the mean of the following data of ages of married men in a certain town.

Ages	15–24	25–34	35–44	45–54	55–64	65–74
No. of Men	33	264	303	214	128	58

NOTES

6. Calculate the mean deviation from the following data. What light does it throw on the social conditions of the community?

   Difference in age between husband and wife:

Difference in Years:	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40
Frequency:	449	705	507	281	109	52	16	4

7. The following figures give the income of 10 persons in rupees. Find the standard deviation.

   114, 115, 123, 120, 110, 130, 119, 118, 116, 115

8. Calculate the mean and standard deviation of the following values of the world's annual gold output millions of pound (in for 20 different years)

   94, 95, 96, 93, 87, 79, 73, 69, 68, 67, 78, 82, 83, 89, 95, 103, 108, 117, 130, 97.
   Also calculate the percentage of cases lying outside the mean at distances $\pm\sigma$, $\pm 2\sigma$, $\pm 3\sigma$ where σ denotes standard deviation.

9. Calculate standard deviation from the following data:

Size of Item	6	7	8	9	10	11	12
Frequency	3	6	9	13	8	5	4

10. From the following information about the accidents on a road in 200 days, calculate the mean number of accidents and the variance of accidents.

No. of Accidents Per Day	0	1	2	3	4	5
No. of Days	46	76	38	25	10	5

11. Calculate the arithmetic mean and the standard deviation from the following data:

Class Interval	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45
Frequency	6	5	15	10	5	4	3	2

12. Calculate the mean and the standard deviations from the following data:

Age Group	Below 20–25	25–30	30–35	35–40	40–45	45–50	50–55	55andabove
No. of Employees	26	44	60	101	109	846	66	10

13. Calculate mean and standard deviation from the following data:

Age Under	10	20	30	40	50	60	70	80
No. of Persons Dying	15	30	53	75	100	110	115	125

14. The marks obtained by the students of class *A* and *B* are given below:

Marks	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45
Class *A*	1	10	20	8	6	3	1	–
Class *B*	5	6	15	10	5	4	2	2

   Calculate mean, median, mode and standard deviation for the distributions. Explain your results regarding composition of the class in respect to intelligence.

15. Explain clearly the ideas implied in using arbitrary working origin and scale for the calculation of the arithmetic mean and standard deviation of frequency distribution.

The values of arithmetic mean and standard deviation of the following frequency distribution of a continuous variable derived from analysis are Rs 135.33 and Rs 9.6 respectively. Find the upper and lower limits of the various classes:

X'	–4	–3	–2	–1	0	+1	+2	+3
f	2	5	8	18	22	13	8	4

16. Obtain the estimating equation by the method of least squares from the following information:

X (*Independent variable*)	*Y* (*Dependent variable*)
2	18
4	12
5	10
6	8
8	7
11	5

17. Find out the coefficient of correlation between the two kinds of assessment of M.A. students' performance.

    (*i*) By adopting Karl Pearson's method

    (*ii*) By the method of least squares

S.N. of Students	*Internal assessment* (*Marks obtained out of 100*)	*External assessment* (*Marks obtained out of 100*)
1	51	49
2	63	72
3	73	74
4	46	44
5	50	58
6	60	66
7	47	50
8	36	30
9	60	35

Also, work out r_2 and interpret the same.

18. Calculate correlation coefficient from the following results:

$$n = 10; \sum X = 140; \sum Y = 150$$
$$\sum (X-10)^2 = 180; \sum (Y-15)^2 = 215$$
$$(X-10)(Y-15) = 60$$

9.12 FURTHER READINGS

NOTES

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 10 INTRODUCTION TO SPSS

NOTES

Structure

10.0 INTRODUCTION

Statistical Package for the Social Sciences or SPSS is a computer program used for statistical analysis. Between 2009 and 2010 the premier vendor for SPSS was named PASW (Predictive Analytics Soft-Ware) Statistics. It was later acquired by IBM and as of January 2010 it became 'SPSS: An IBM Company'. The features of SPSS incorporate modules for statistical data analysis, including descriptive statistics such as plots, frequencies, charts and lists, as well as sophisticated inferential and multivariate statistical procedures, such as ANalysis Of VAriance (ANOVA), factor analysis, cluster analysis and categorical data analysis. In addition to statistical analysis, data management and data documentation are unique features of the base software. The various features of SPSS are easily accessible through pull-down menus or can also be programmed using a proprietary 4GL (Fourth-Generation Language) command syntax language. SPSS is a modular product hence it requires the Base System module to run.

SPSS Statistics 17.0 is a comprehensive system for analysing data based on the GUI. SPSS Statistics can obtain data from almost any type of file and use them to produce tabulated reports, charts, plots of distributions and trends,

NOTES

descriptive statistics and complex statistical analyses. In addition, below the menus and dialog boxes, SPSS Statistics uses a command language for data analysis.

The syntax editor in SPSS Statistics 17.0 has been entirely redesigned including features, such as auto completion, color coding, bookmarks and breakpoints. Analysing data using SPSS Statistics is simple and easy. The Data Editor provides a convenient, spreadsheet like method for creating and editing data files. The Data Editor window opens automatically when the user starts SPSS. The dimensions of the data file are determined by the number of cases and variables. You can enter data in any cell. Variable names can be up to 64 bytes long and the first character must be a letter or one of the characters @, # or $. Subsequent characters can be any combination of letters, numbers, non-punctuation characters and a period. The user can also specify the level of measurement as scale (numeric data on an interval or ratio scale), ordinal or nominal as per his/her requirements. Nominal and ordinal data can be either string (Alphanumeric) or numeric.

In this unit, you will study about the SPSS software, basic structure of an SPSS Data File, using Data Editor, reading Spreadsheet Data, reading a Database, and reading Text Data.

10.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the significance of SPSS software
- Discuss the basic structure of an SPSS Data File
- Use and work with Data Editor
- Read the Spreadsheet Data
- Analyse and read a Database file
- Elaborate on how Text Data is read

10.2 ABOUT SPSS

SPSS is abbreviated term for Statistical Package for the Social Sciences and is used for data management and analysis. This program is used on computers for statistical analysis in social science by government, market researchers, education researchers, health researchers and survey companies. The statistical package SPSS is used to perform quantitative research in social science because it is easy to use. The SPSS Data Editor is very valuable and is specifically designed for performing statistical tests, such as correlation, regression, *t*-test, hypotheses, Chi-square and ANalysis Of VAriance or ANOVA. It also helps a researcher to make useful data entries, find frequency counts, sort and rearrange data, etc.

The SPSS features available with the software package can be accessed with the help of pull- down menus or can be programmed using a licensed 4GL (Fourth-Generation Language) command syntax language. The advantage of command syntax programming language is that it helps in data reproducibility, simplifying repetitive tasks, performing complex data manipulations and analyses. In addition, the user can program specific syntax for some complex applications which are not available in the predefined menu structure. The command syntax can also be generated by pull-down menu interface and can be displayed in the output. To syntax can be made visible to the user by changing the default settings. It can also be pasted into a syntax file with the help of 'paste' button which is available in each menu.

SPSS can read and write data from ASCII (American Standard Code for Information Interchange) text files including hierarchical files, other statistics packages, spreadsheets and databases. SPSS can also be used to read and write to external relational database tables using ODBC (Open DataBase Connectivity) and SQL (Sequential Query Language). Statistical output is in the licensed file format with the file extension name as **.spv** which supports pivot tables. The output can be exported to Microsoft Word and can be acquired as data, as text, PDF, XLS, HTML, XML, SPSS dataset or in the graphic image formats (JPEG, PNG, BMP and EMF).

The SPSS is based on Graphical User Interface (GUI) which supports the two data editor views, the Data View and the Variable View. The user can toggle between the two views just by selecting one of the two tabs that appear in the bottom left of the SPSS window and clicking on it. The 'Data View' exhibits a view in the form of a spreadsheet as the cases (rows) and variables (columns). Only two data types can be defined in SPSS Statistics, i.e., the numeric data type and the text or 'string' data type. All data processing processes appears in sequence case-by-case through the file. You can match the files on the basis of one-to-one and one-to-many, but not many-to-many. In SPSS, the data cells simply hold numbers or text. You cannot store the formulas in these cells. The 'Variable View' exhibits the metadata dictionary in which each row represents a variable to display the variable name, variable label, value label(s), print width, measurement type and other associated characteristics. In both views, you can manually edit the cells, define file structure and do data entry without using the command syntax for smaller datasets. Large datasets, such as statistical surveys are created using data entry software or entered by scanning using Optical Character Recognition (OCR) and Optical Mark Recognition (OMR) software. Using a 'macro' language command language subroutines can be written. A Python programmability extension is used to access the information in the data dictionary and dynamically build command syntax programs. Figure 10.1 illustrates the initial Windows Dialog Box for SPSS.

NOTES

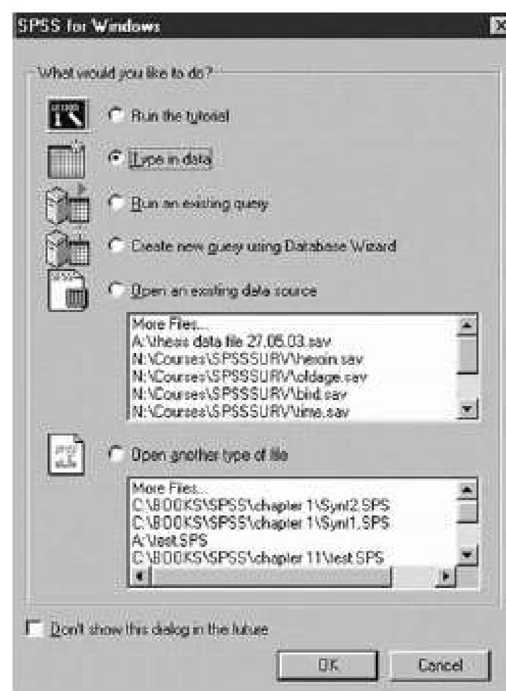


Fig. 10.1 *Initial Windows Dialog Box for SPSS*

10.2.1 Why Syntax?

The syntax is basically a text file where you can add comments and SPSS commands. The reasons for why everyone should use the syntax function are:

- It is a way of documenting and archiving everything you have done with the data material.
- It is easy to repeat parts or all of the analysis.
- Other people involved in the data material can easily understand what you have done and how you have done it.
- It saves an enormous amount of time

10.2.2 Log-In to SPSS

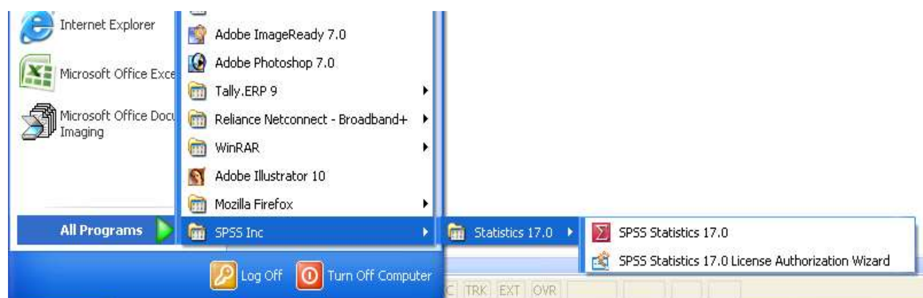
There are two ways to start the SPSS program. One is to simply click on the SPSS icon shown on your desktop. If you cannot find the icon, you can click Start on the bottom of your screen, then Program Files, and then SPSS. Or if you are not sure whether the computer you are using has SPSS, click Start, then Find, then Files or Folders, then type 'SPSS'. When the SPSS window starts, a dialogue box will pop up as shown below. You can either start and type in new data, or open an existing file.

NOTES



10.3 WORKING WITH SPSS

Install SPSS on your computer and create a **shortcut** menu on your desktop and to directly start the package by clicking on the SPSS icon. Alternatively, you can go to the **Start → All Programs → SPSS Inc → Statistics 17.0 → SPSS Statistics 17.0** as shown:



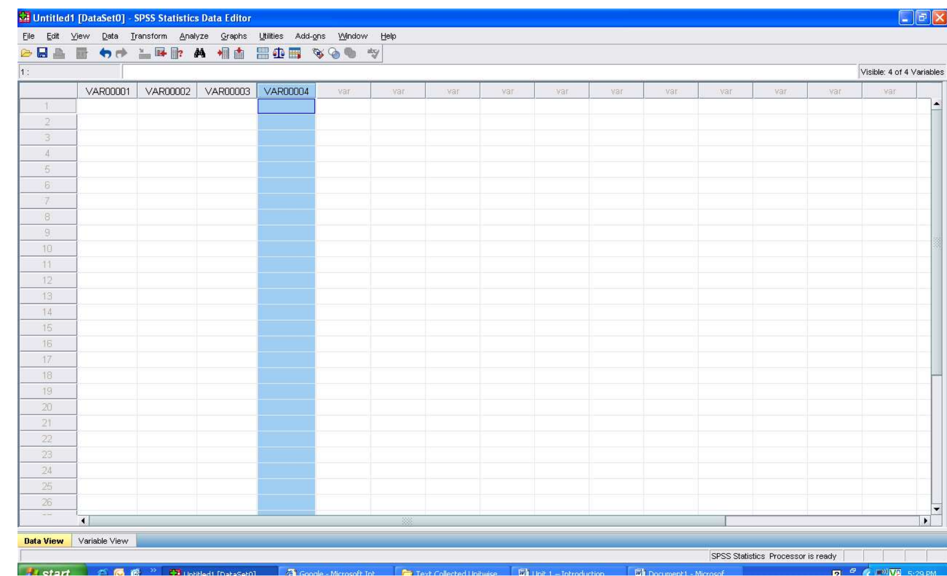
When you click on the SPSS Statistics 17.0, the following screen will appear to start SPSS program:



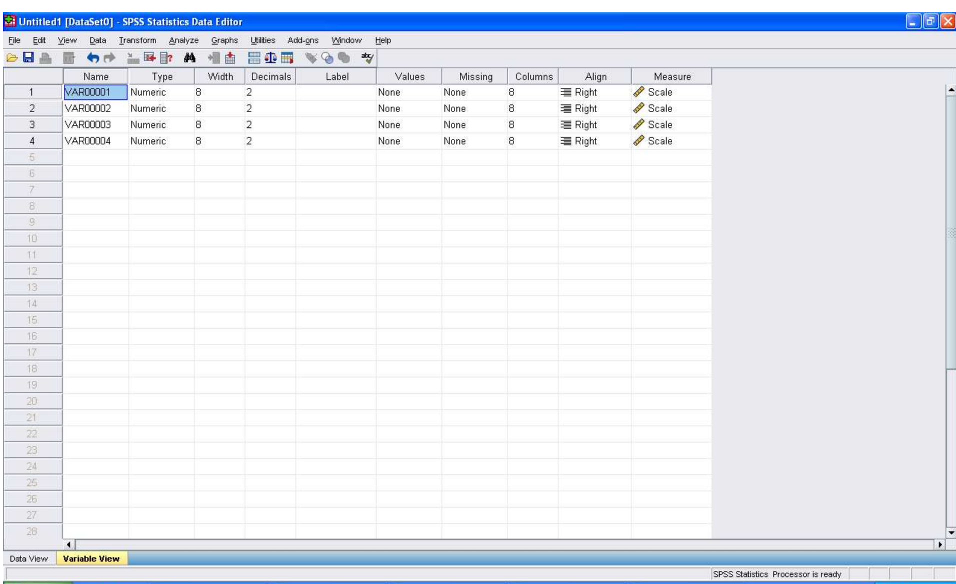
NOTES

The SPSS Data Editor has two main 'Views'. User can enter data in the Data View while in the Variable View, user can select the name, type, maximum number of letters per cell ('Width'), number of decimal points, label, width of cell ('Columns'), alignment within the cell ('Align') and whether or not the variable is nominal, ordinal or 'Scale' ('Measure'). The user can also categorize entries into labels (in the 'Values' column) and mark entries as invalid (in the 'Missing' column) in the Variable View. Using SPSS for Windows the user can perform almost any statistics calculation in combination with pointing and clicking on the menus and various specific interactive dialog boxes.

Data View



Variable View



NOTES

10.3.1 SPSS Statistics 17.0

SPSS Statistics 17.0 is a comprehensive system for analysing data based on the GUI. SPSS Statistics can acquire data from almost any form of file and use them to create tabulated reports, charts, plots of distributions and trends, descriptive statistics and complex statistical analyses. SPSS Statistics Base 17.0 provides examples in the help system which is automatically installed with the software. In addition, below the menus and dialog boxes, SPSS Statistics uses a command language for data analysis.

SPSS Statistics has a powerful statistical analysis and data management system in a graphical environment. It also has descriptive menus and simple dialog boxes which help the users to accomplish the task just by pointing and clicking the mouse. In addition to the simple point-and-click interface for statistical analysis, SPSS Statistics provides the following features:

- **Data Editor:** The Data Editor is similar to multipurpose spreadsheet system and is used to define, enter, edit and display data.
- **Viewer:** The Viewer helps to browse the results, show and hide selective outputs, modify the display order results, shift presentation quality tables and charts to and from other applications.
- **Multidimensional Pivot Tables:** The multidimensional pivot tables display the output results in the form which look alive. Users can explore tables by rearranging rows, columns and layers. It is also easy to compare the groups. It is done by splitting the table so that only one group is displayed at a time.
- **High Resolution Graphics:** High resolution, full color pie charts, bar charts, histograms, scatter plots, 3D graphics, etc., are built-in standard features.
- **Database Access:** The user can directly recover information from databases by using the Database Wizard omitting the complex SQL queries.
- **Data Transformations:** Transformation features help to find the data organized for analysis. You can also subset data and files to combine categories, add, aggregate, merge, split, transpose and much more.
- **Online Help:** A comprehensive abstract of context sensitive Help topics are available in dialog boxes to guide the users while performing specific tasks, pop-up definitions in pivot table results, explaining statistical terms. The Statistics Coach helps the users to find the required procedures whereas case studies provide hands-on examples for using statistical procedures and to interpret the results.
- **Command Language:** Most of the tasks in SPSS Statistics are completed with the help of simple point-and-click actions. It also provides a powerful command language which permits the user to save and automate

various common tasks. The command language also provides several functionalities which are not available in the menus and dialog boxes. Complete command syntax documentation is incorporated into the overall Help system.

NOTES

10.3.2 What's New in SPSS Statistics Version 17.0?

The following are the enhanced features available in the new version of SPSS Statistics 17.0:

New Syntax Editor: The syntax editor in SPSS Statistics 17.0 has been entirely redesigned including features, such as auto completion, color coding, bookmarks and breakpoints. Auto completion feature provides the user a list of valid command names, subcommands and keywords hence the user spends less time referring to syntax charts. Color coding feature permits the user to spot unrecognized terms and some common syntactical errors quickly. Bookmarks feature permits the user to speedily navigate huge command syntax files. Breakpoints feature permits the user to stop execution at specific points for inspecting data or output prior to proceeding.

- **Custom Dialog Builder:** The Custom dialog builder permits the user to create and manage custom dialogs for creating command syntax.
- **Multiple Language Support:** In addition to the capability to modify the output language the user can now modify the user interface language.
- **Codebook:** The Codebook method accounts the dictionary information, such as variable names, variable labels, value labels, missing values and summary statistics for all or specific variables and manifold response sets in the active dataset. For nominal and ordinal variables, and manifold response sets, summary statistics include counts and percent. For scale variables, summary statistics include mean, standard deviation and quartiles.
- **Nearest Neighbor Analysis:** Nearest neighbor analysis is a technique for categorizing cases based on their similarity to other cases. In machine learning, it was used to identify patterns of data without involving an accurate match to any stored patterns or cases. Similar cases are close to each other and dissimilar cases are isolated from each other. Thus, the distance between two cases is a measure of their dissimilarity.
- **Multiple Imputation:** The Multiple Imputation method executes multiple imputation of missing data values. The dataset having missing values give outputs as one or more datasets in which missing values are substituted with plausible estimates. The pooled results obtained when other procedures are run. This technique also summarizes missing values in the working dataset. This feature is available in the Missing Values add-on option.

- **RFM Analysis:** RFM analysis is the abbreviated form of Recency, Frequency, Monetary Analysis. This method is used to recognize existing customers who are most probable to respond to a new offer and is frequently used in direct marketing. This feature is available in the EZ RFM add-on option. The fundamental principle of RFM analysis is for customers who have purchased recently, have made more purchases and are more likely to respond to your offering than other customers who have purchased less recently, less often and in smaller amounts.

Basically, RFM analysis uses information about customers' past behavior that is easily tracked and readily available. **Recency** is how long ago the customer last made a purchase. **Frequency** is how many purchases the customer has made (sometimes within a specified time period, such as average number of purchases per year). **Monetary** is total amount spent by the customer (sometimes within a specified time period).

- **Categorical Regression Enhancements:** Categorical Regression has been enhanced and included regularization and resampling techniques for accurately assessing and improving predictions. Jointly, these new methods feasibly create state-of-the-art models even for high volume data where there are more variables than observations. This feature is available in the Categories add-on option.
- **Graph-Board:** Graph-Board are visualizations which include graphs, charts and plots created using a visualization template. SPSS Statistics 17.0 provides built-in new visualization templates which are effectively custom visualization types.
- **Exporting Output:** The following output export format options and control over exported contents are available in SPSS Statistics version 17.0:
  - o To wrap or shrink wide table in Word documents.
  - o To create new worksheets or append data to existing worksheets in an Excel workbook.
  - o To save output export specifications in the form of command syntax with the **OUTPUT EXPORT** command. All the features for exporting output in the Export Output dialog are available in command syntax.
  - o The Output Management System (OMS) supports the additional output formats, such as Word, Excel and PDF.
- **Shift Values:** Shift Values generates new variables that hold the values of existing variables from preceding or subsequent cases.
- **Aggregate Enhancements:** This feature allows the user to use the aggregate method without specifying a break variable.
- **Median Function:** A median function is available for computing the median value across selected variables for each case.

NOTES

10.4 WINDOWS IN SPSS STATISTICS

The following are the different types of windows in SPSS Statistics:

- **Data Editor:** The Data Editor displays the contents of the data file. You can create new data files or modify existing data files using the Data Editor. When you open more than one data file then there is a separate Data Editor window for each opened data file.
- **Viewer:** The Viewer displays all statistical results, tables and charts. The user can edit the output and save it for later use. A Viewer window opens automatically the first time the user runs a procedure to generate output.
- **Pivot Table Editor:** The Pivot Table Editor modifies the output in various ways that is displayed in pivot tables. The user can edit text, swap data in rows and columns, add color, create multidimensional tables, and hide and show selective results.
- **Chart Editor:** The high resolution charts and plots can be modified in chart windows. The user can change the colors, select different font types or sizes, switch the horizontal and vertical axes, rotate 3D scatter plots and even change the chart type.
- **Text Output Editor:** Text output which is not displayed in pivot tables can be modified using the Text Output Editor. The user can edit the output and modify font characteristics, such as type, style, color and size.
- **Syntax Editor:** The user can paste the dialog box choices into a syntax window, where the selections appear in the form of command syntax. Now edit the command syntax to use special features that are not available through dialog boxes. The user can also save these commands in a file for use in subsequent sessions.

10.5 USING DATA EDITOR

Unlike commonly-used Microsoft Office applications such as Word and Excel, SPSS has many windows. It can be quite confusing in the beginning; you will get used to it as you work along. The spreadsheet window is called the **Data Editor.** You can also open an output viewer, syntax editor and script editor window from the 'File' menu through 'New' or 'Open'. Later, we will activate the Chart Editor and Pivot Table Editor. The most important windows are the **Data Editor** and the **SPSS Viewer**. The Data Editor displays data and allows data manipulation and analysis. The **SPSS Viewer** displays output and keeps log of changes in the program. The **Syntax Window** displays the command instructions; it helps keep track of analysis and perform automated tasks. The **Chart Editor** and **Pivot Table Editor** are for editing charts and tables. The **Script Editor** is mostly for making specialized formats of table output. The Syntax window and Script Editor are for more

NOTES

experienced users. All the windows of more complicated tasks will be shown only when you actually activate them. You can toggle between these windows by clicking on the taskbar at the bottom of your screen or by selecting a window from the Window pull-down menu at the top. The pull-down (or drop-down) menus at the top of your screen are similar to Office applications. There are common categories like File, Edit, View, Windows and Help, which you can figure out by looking at the names. Specific to SPSS are four categories, Data, Transform, Analyze (Statistics in earlier versions) and Graphs in the Data Editor Window. It's slightly changed in the Output Viewer window; Data and Transform are replaced by Insert and Format. Some of the commands have icons on the toolbar which provide shortcuts. If you place the cursor on an icon, its name should appear. The Data Editor provides a convenient, spreadsheet like method for creating and editing data files. The Data Editor window opens automatically when the user starts SPSS. The Data Editor provides the following two views of the data:

- **Data View:** This view displays the actual data values or defined value labels.
- **Variable View:** This view displays variable definition information including defined variable and value labels, Data Type (for example, String, Date or Numeric), Measurement Level (Nominal, Ordinal or Scale) and user defined missing values.

  In both views, the user can add, change and delete information contained in the data file.

  Data View: Many of the features of Data View are similar to the features that are found in spreadsheet applications. The following are the significant differences:

  - **Rows are Cases:** Each row represents a case or an observation. For example, each individual respondent to a questionnaire is a case.
  - **Columns are Variables:** Each column represents a variable or characteristic that is being measured. For example, each item on a questionnaire is a variable.
  - **Cells contain Values:** Each cell contains a single value of a variable for a case. Where the case and the variable intersect is called the cell. Cells contain only data values. Unlike spreadsheet programs, cells in the Data Editor cannot contain formulas.
  - **The Data File is Rectangular:** The dimensions of the data file are determined by the number of cases and variables. You can enter data in any cell. If you enter data in a cell outside the boundaries of the defined data file, the data rectangle is extended to include any rows and/or columns between that cell and the file boundaries. There are no 'empty' cells within the boundaries of the data file. For numeric variables, blank cells are converted to the system missing value. For string variables, a blank is considered a valid value.

NOTES

Variable View: Variable View contains descriptions of the attributes of each variable in the data file. In Variable View, the variables can be added or deleted and attributes of variables can be modified including the following attributes:

- Rows are Variables
- Columns are Variable Attributes
- Variable Name
- Data Type
- Number of Digits or Characters
- Number of Decimal Places
- Descriptive Variable and Value Labels
- User Defined Missing Values
- Column Width
- Measurement Level

All of these attributes are saved when the user saves the data file. In addition to defining variable properties in Variable View, the following are two other methods for defining variable properties:

- **The Copy Data Properties Wizard:** This wizard provides the ability to use an external SPSS Statistics data file or another dataset that is available in the current session as a template for defining file and variable properties in the active dataset. The user can also use variables in the active dataset as templates for other variables in the active dataset. Copy Data Properties is available on the Data menu in the Data Editor window.
- **Define Variable Properties:** This is also available on the Data menu in the Data Editor window. It scans the specific data and lists all unique data values for any selected variables, identifies unlabeled values and provides an auto label feature. This method is used for categorical variables that use numeric codes to represent categories, for example, '0 = Male', '1 = Female'.

Check Your Progress

1. Define SPSS.
2. What are the major reasons for using the syntax function in SPSS?
3. How do you log-in to the SPSS?
4. State the features of SPSS.
5. What are the various types of windows in SPSS?
6. What do you mean by variable view in data editor?

10.6 READING SPREADSHEET DATA

10.6.1 The Data View in a Spreadsheet

The Data Editor consists of two windows. By default the Data View, which allows the data to be entered and viewed, is shown (Refer Figure 10.2). The other window is the Variable View, which allows the types of variables to be specified and viewed. The user can toggle between the windows by clicking on the appropriate tabs on the bottom left of the screen. Data values can be entered in the Data View spreadsheet. For most analysis SPSS assumes that rows represent cases and columns variables. For example, in Figure 10.2 some of five available variable values have been entered for twenty subjects. By default SPSS aligns numerical data entries to the right-hand side of the cells and text (string) entries to the left-hand side. Here variables sex, age, extrover, and car take numerical values while the variable make takes string values. By default SPSS uses a period/full stop to indicate missing numerical values. String variable cells are simply left empty. Here, for example, the data for variables extrover, car, and make have not yet been typed in for the 20 subjects so the respective values appear as missing.

data11.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

1 : sex 1

	sex	age	extrover	car	make
1	1	55	40	46	Ford
2	1	43	45	79	Rolls Royce
3	0	57	52	33	Citroen
4	1	26	62	63	Renault
5	0	22	31	20	
6	0	32	28	18	
7	0	26	2	11	
8	1	29	83	97	
9	1	40	55	63	
10	0	30	32	46	
11	0	34	47	.	
12	1	44	45	.	
13	1	49	60	.	
14	1	22	.	.	
15	0	34	.	.	
16	1	47	.	.	
17	0	48	.	.	
18	0	48	.	.	
19	1	22	.	.	
20	0	24	.	.	

Data View / Variable View

SPSS Processor is ready

Fig. 10.2 Data View Window of the Data Editor

The appearance of the Data View spreadsheet is controlled by the View drop-down menu. This can be used to change the font in the cells, remove lines, and make value labels visible. When labels have been assigned to checking the category codes of a categorical variable, these can be displayed by labels are

NOTES

visible, highlighting a cell produces a button with a downward Value Labels (or by selecting on the toolbar). Once the category arrow on the right-hand side of the cell. Clicking on this arrow produces a drop-down list with all the available category labels for the variable. Clicking on any of these labels results in the respective category and label being inserted in the cell. This feature is useful for editing the data.

10.6.2 The Variable View in a Spreadsheet

The Variable View spreadsheet serves to define the variables (Refer Figure 10.3). Each variable definition occupies a row of this spreadsheet. As soon as data is entered under a column in the Data View, the default name of the column occupies a row in the Variable View.

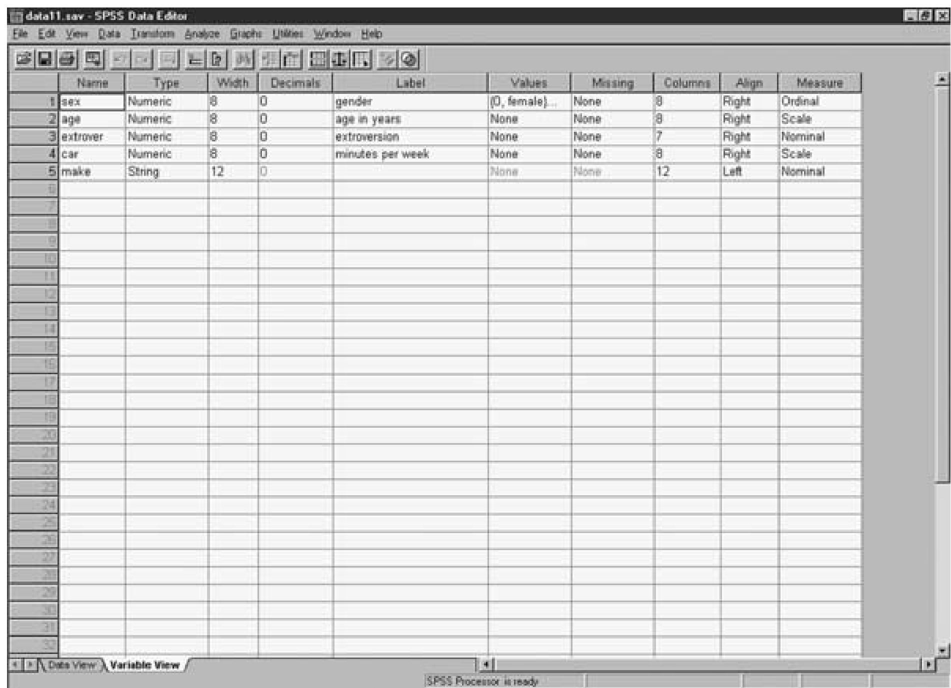


Fig. 10.3 Variable View Window of the Data Editor

There are 10 characteristics to be specified under the columns of the Variable View (Refer Figure 10.3):

1. **Name:** The chosen variable name. This can be up to eight alphanumeric characters but must begin with a letter. While the Underscore (_) is allowed, Hyphens (-), Ampersands (&), and Spaces( )cannot be used. Variable names are not case sensitive.
2. **Type:** The type of data. SPSS provides a default variable type once variable values have been entered in a column of the Data View. The type can be changed by highlighting the respective entry in the second column of the Variable View and clicking the three-period symbol

(…) appearing on the right-hand side of the cell. This results in the Variable Type box being opened, which offers a number of types of data including various formats for numerical data, dates, or currencies.

3. **Width:** The width of the actual data entries. The default width of numerical variable entries is eight. The width can be increased or decreased by highlighting the respective cell in the third column and employing the upward or downward arrows appearing on the

   Right-hand side of the cell or by simply typing a new number in the cell.
4. **Decimals:** The number of digits to the right of the decimal place to be displayed for data entries. This is not relevant for string data and for such variables the entry under the fourth column is given as a zero. The value can be altered in the same way as the value of Width.
5. **Label:** A label attached to the variable name. In contrast to the variable name, this is not confined to eight characters and spaces can be used. It is generally a good idea to assign variable labels. They are helpful for reminding users of the meaning of variables (placing the cursor over the variable name in the Data View will make the variable label appear) and can be displayed in the output from statistical analyses.
6. **Values:** These are the labels attached to category codes. For categorical variables, an integer code should be assigned to each category and the variable defined to be of type 'Numeric'. When this has been done, clicking on the respective cell under the sixth column of the Variable View makes the three-period symbol appear, and clicking this opens the Value Labels dialogue box, which in turn allows assign-ment of labels to category codes. For example, our data set included a categorical variable sex indicating the gender of the subject. Clicking the three-period symbol opens the dialogue box shown in Display 1.4 where numerical code '0' was declared to represent females and code '1' males.
7. **Missing:** Missing value codes. SPSS recognizes the period symbol as indicating a missing value. If other codes have been used (e.g., 99, 999) these have to be declared to represent missing values by highlighting the respective cell in the seventh column, clicking the three-period symbol and filling in the resulting Missing Values dia-logue box accordingly.
8. **Columns:** Width of the variable column in the Data View. The default cell width for numerical variables is eight. Note that when the Width value is larger than the Columns value, only part of the data entry might be seen in the Data View. The cell width can be changed in the same way as the width of the data entries or simply by dragging the

NOTES

relevant column boundary. Place cursor on right-hand bound-ary of the title of the column to be resized. When the cursor changes into a vertical line with a right and left arrow, drag the cursor to the right or left to increase or decrease the column width.

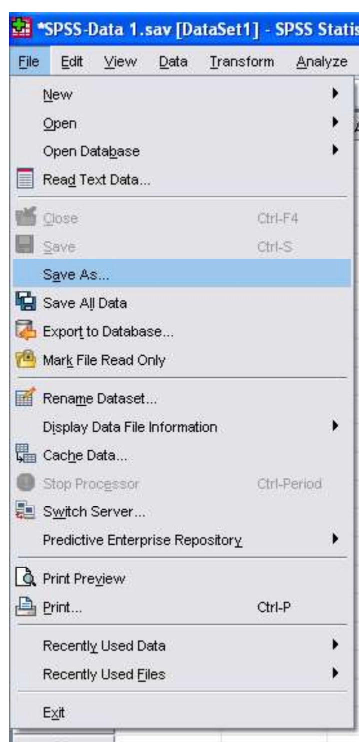
9. **Align:** Alignment of variable entries. The SPSS default is to align numerical variables to the right-hand side of a cell and string variables to the left. It is generally helpful to adhere to this default; but if necessary, alignment can be changed by highlighting the relevant cell in the ninth column and choosing an option from the drop-down list.
10. **Measure:** Measurement scale of the variable. The default chosen by SPSS depends on the data type. For example, for variables of type 'Numeric', the default measurement scale is a continuous or interval scale (referred to by SPSS as 'Scale'). For variables of type 'String', the default is a nominal scale. The third option, 'Ordinal', is for categorical variables with ordered categories but is not used by default. It is good practice to assign each variable the highest appropriate measurement scale ('Scale' > 'Ordinal' > 'Nominal') since this has implications for the statistical methods that are applicable. The default setting can be changed by highlighting the respective cell in the tenth column and choosing an appropriate option from the drop-down list.

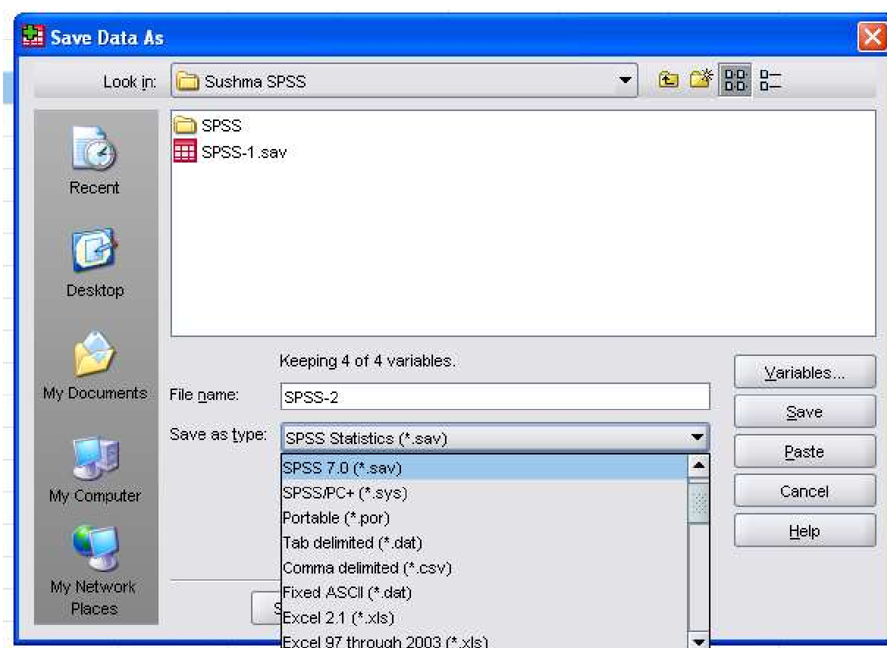
10.7 STORING AND RETRIEVING DATA FILES

Storing and retrieving data files are carried out using the drop down menu available on the menu bar once you select a File. In the Data Editor a data file can be saved by using the commands **File → Save or Save As…**as shown in the screen. The first time you save the file you have to use Save As command for saving the file at the specified location in the specific folder. When the file already exits, then you can use Save command to save the current changes or modifications.

In the standard Windows method the Save command will save the data file under its current name, overwriting an existing file or prompting for a name otherwise whereas the Save As command always opens the Save Data As dialog window where the directory, file name and type have to be specified by the user. SPSS supports a number of data formats. SPSS data files are given the extension **.sav** as shown below. Other formats are also available for saving the file, such as ASCII text is saved as **.dat**, Excel file is saved as **.xls** and dBASE file is saved as **.dbf**.

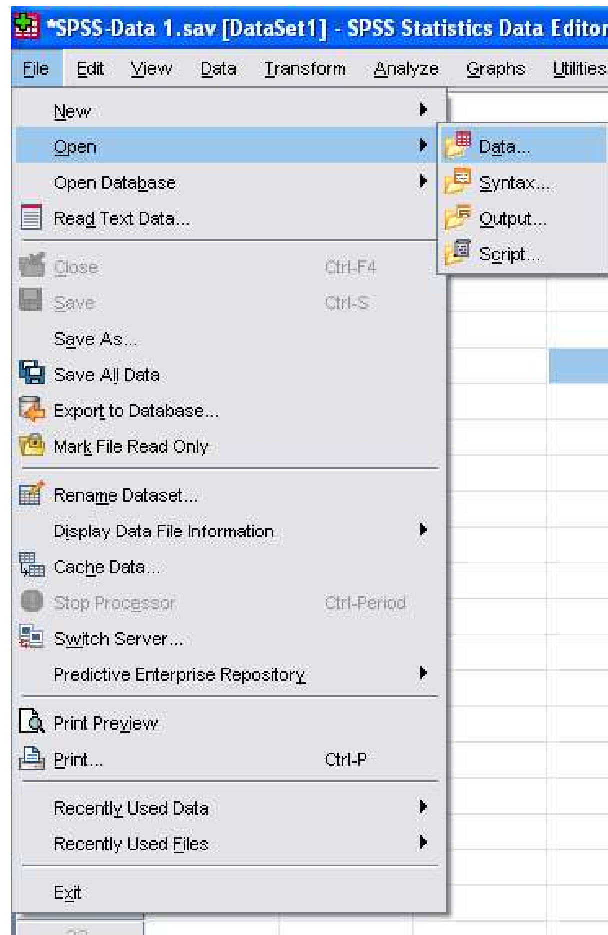
NOTES


*SPSS-Data 1.sav [DataSet1] - SPSS Statis
File Edit View Data Transform Analyze
New
Open
Open Database
Read Text Data...
Close Ctrl-F4
Save Ctrl-S
Save As...
Save All Data
Export to Database...
Mark File Read Only
Rename Dataset...
Display Data File Information
Cache Data...
Stop Processor Ctrl-Period
Switch Server...
Predictive Enterprise Repository
Print Preview
Print... Ctrl-P
Recently Used Data
Recently Used Files
Exit


Save Data As
Look in: Sushma SPSS
SPSS
SPSS-1.sav
Recent
Desktop
My Documents
My Computer
My Network Places
Keeping 4 of 4 variables.
File name: SPSS-2
Save as type: SPSS Statistics (*.sav)
SPSS 7.0 (*.sav)
SPSS/PC+ (*.sys)
Portable (*.por)
Tab delimited (*.dat)
Comma delimited (*.csv)
Fixed ASCII (*.dat)
Excel 2.1 (*.xls)
Excel 97 through 2003 (*.xls)
Variables...
Save
Paste
Cancel
Help

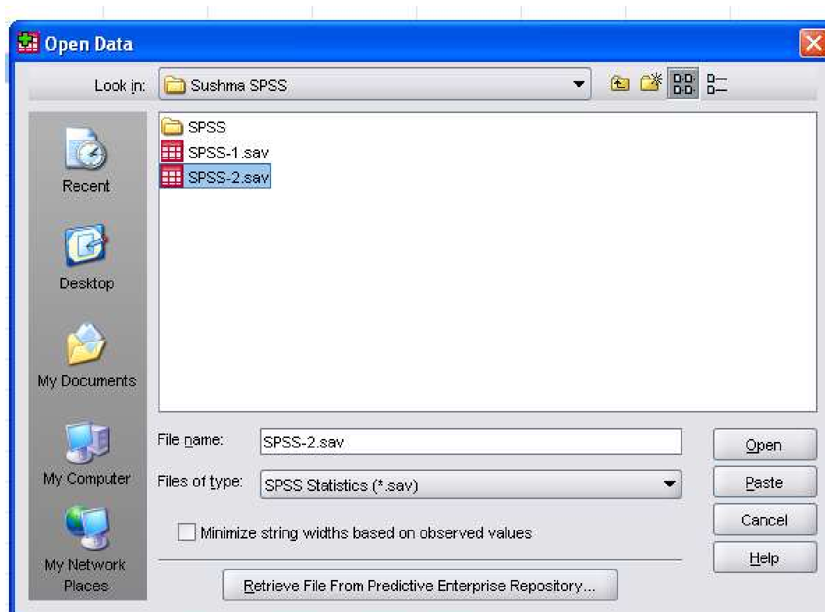
NOTES

To open an existing SPSS data file, from the main menu select **File → Open → Data…** as shown in the screen.

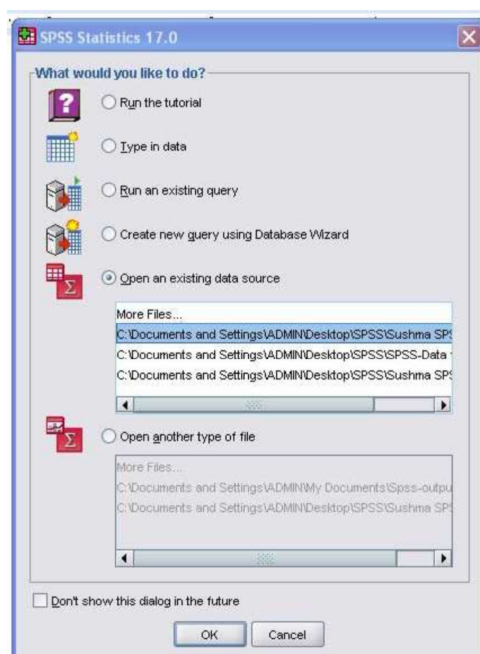


This will open the Open File Dialog Box from which the appropriate file can be selected in the standard way as shown below. Recently used files are also accessible by placing the cursor over Recently Used Data on the File drop-down menu and double clicking on the required file.

NOTES



In addition, files can be opened when first starting SPSS by checking **Open an existing data source** on the initial dialog box as shown below.



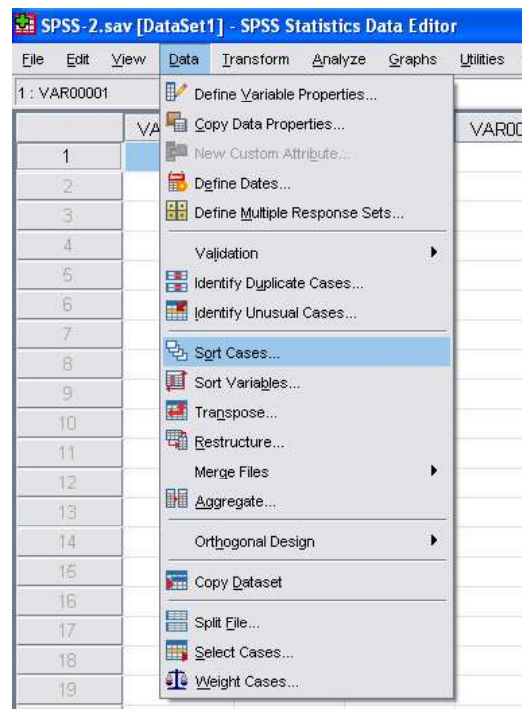
SPSS can also import data files in other than SPSS format. A list of data formats is provided by selecting the down arrow next to the Files of type field. There are a number of formats, such as spreadsheet or Excel as **.xls**, database or dBase as **.dbf**, and ACSII text as **.txt** and **.dat** which can be imported by selecting a particular file extension. A dialog box will appear that asks for information relevant to the format.

NOTES

10.7.1 Data File Handling

The data file as displayed in the Data View spreadsheet is not always organized in the proper format for a specific use. The Data drop down menu provides methods for reorganizing the structure of a data file. The first four command options from the Data drop down menu are concerned with editing or moving within the Data View spreadsheet. Date formats can be defined or variables or cases inserted. The following set of methods set a limit for the format of a data file to be changed/altered:

Sort Cases…: It opens a dialog box that allows sorting of cases (rows) in the spreadsheet according to the values of one or more variables. Cases can be arranged in ascending or descending order. When several sorting variables are in use then cases will be sorted by each variable within categories of the prior variable on the Sort by list. Sorting is very useful for generating graphics.

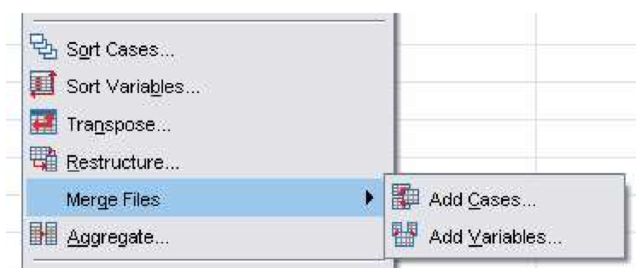


Sort Variables…: It sorts Data Values within a Variable Column. To sort numerical data in a matrix, click on 'Data' and then click on 'Sort Cases'. Click on the variable that you want to sort as the primary variable. This will move the variable title into the **Sort by:** column. The Sort Order can be either ascending or descending. This command is used to sort the data and automatically rearrange all the other variables.

Transpose…: It opens a dialog for swapping the rows and columns of the data spreadsheet. The Variable(s) list contains the columns to be transposed into rows and an ID variable can be supplied as the Name Variable to name the columns of the new transposed spreadsheet. This command is used when procedures generally used on the columns of the spreadsheet are to be applied to the rows.

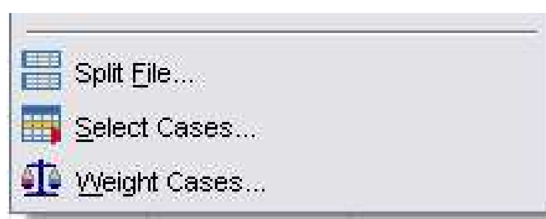
Restructure…: It calls the **Restructure Data Wizard**, which includes a series of dialog boxes for converting data spreadsheets between what are known as 'long' and 'wide' formats. These formats are relevant in the analysis of repeated measures.

Merge Files: It allows either **Add Cases…** or **Add Variables…** to an existing data file as shown below. A dialog box is displayed that permits opening a second data file. This file can either contain the same variables but different cases (to add cases) or different variables but the same cases (to add variables). These commands are useful at the building phase of a database project and suggest wide ranging options to combine data files.



Aggregate…: It merges groups of rows (cases) into single abstract row and creates a new aggregated data file. The grouping variables are provided under the Break Variable(s) list of the Aggregate Data dialog box and the variables to be aggregated under the Aggregate Variable(s) list. The **Function…** sub- dialog box allows for the aggregation function of each variable to be selected from a number of possibilities (mean, median, value of first case, number of cases, etc.). This command is helpful when the data structure is of hierarchical nature.

Finally, the **Split File…**, **Select Cases…** and **Weight Cases…** procedures allow using the data file in a specific format without changing its appearance in the Data View spreadsheet. These commands are frequently used in data analysis. The Weight Cases… command (or from the toolbar) is typically used in connection with categorical data — it internally replicates rows according to the values of a Frequency Variable. It is useful when data is provided in the form of a cross tabulation.



10.8 READING TEXT DATA IN SPSS

This module introduces the reading of raw data files into SPSS. It does not show how to read all possible data formats, but aims to show how to read many common file formats such as:

NOTES

1. Comma Delimited Data, Inline

The example below illustrates how you can read comma delimited data inline. Inline data refers to data that is included as part of a syntax program between **BEGIN DATA** and **END DATA** commands. This is a very convenient way to read data when your data file is small and or when you want to be sure to keep your data and the program for reading your data together.

In the **DATA LIST LIST** command, the **DATA LIST** command indicates that we are going to read a data file. The keyword **LIST** (the second one) indicates that there is one observation per line of data and that each variable is separated by spaces and or commas. Following the / is the list of the variable names. Since **MAKE** is a string (Alphanumeric) variable, we use the **(A15)** to tell SPSS that **MAKE** is a string variable (with a length of 15). The data appear between the **BEGIN DATA** and **END DATA** commands, and the **LIST** command is used to see if the data were read correctly.

```
DATA LIST LIST
 / MAKE (A15) MPG WEIGHT PRICE
BEGIN DATA
"AMC Concord", 22,2930,4099
"AMC Pacer", 17,3350,4749
"AMC Spirit", 22,2640,3799
"Buick Century", 20, 3250,4816
"Buick Electra", 15,4080,7827
END DATA
LIST
```

From the output below, it looks like the data were read properly. You might wonder if the quotations marks around the makes of car were needed. SPSS treats commas and spaces as valid delimiters (separators) in this mode of input, so the quotes are necessary. Otherwise, SPSS would think that AMC was the **MAKE** and Concord was the **MPG** for the first observation.

```
MAKE MPG WEIGHT PRICE
AMC Concord 22.00 2930.00 4099.00
AMC Pacer 17.00 3350.00 4749.00
AMC Spirit 22.00 2640.00 3799.00
Buick Century 20.00 3250.00 4816.00
Buick Electra 15.00 4080.00 7827.00
```

2. Space Delimited Data, Inline

Reading space delimited data is basically the same as reading a comma delimited file. As you see below, the program is identical to the one above, but the data are separated by one or more spaces instead of commas. We will omit showing the output from this command.

```
DATA LIST LIST
 / MAKE (A15) MPG WEIGHT PRICE
```

NOTES

```
BEGIN DATA
“AMC Concord” 22 2930 4099
“AMC Pacer” 17 3350 4749
“AMC Spirit” 22 2640 3799
“Buick Century” 20 3250 4816
“Buick Electra” 15 4080 7827
END DATA
LIST
```

3. Comma Delimited Data from an External File

When you have a larger file, it is often more convenient to read the data from an external data file. The syntax of reading data from an external data file is not much more complicated than reading from inline data. The example below shows that you use **file=”auto1.txt”** to tell SPSS that you want to read the data from the file called **auto1.txt**. When you run this, change **auto1.txt** to be the name of your raw data file (for example, **”c:mydata.txt”**). We will omit showing the output of these commands.

```
DATA LIST LIST FILE=”auto1.txt”
 / MAKE (A15) MPG WEIGHT PRICE
LIST
```

4. Space Delimited Data from an External File

As we saw with the inline data, reading a space delimited data file is the same as reading a comma delimited file. As we would expect, the commands below are identical to the ones above, except that it reads from **auto2.txt**, which is a space delimited file.

```
DATA LIST LIST FILE=”auto2.txt”
 / MAKE (A15) MPG WEIGHT PRICE
LIST
```

5. Fixed Format Data from an External File

The data can be read in a fixed format external data file. Consider the data file shown below.

```
AMC Concord 22 2930 4099
AMC Pacer 17 3350 4749
AMC Spirit 22 2640 3799
Buick Century 20 3250 4816
Buick Electra 15 4080 7827
```

The data are in fixed format because the variables occupy the same column positions throughout the entire data file. For example, **MAKE** is always in columns 1-13, **MPG** is always in columns 15-16, **WEIGHT** is always in columns 18-12, and **PRICE** is always in columns 23-26.

NOTES

Assume that these data were stored in a file called **auto3.txt**. The commands below show how you can read this data as a fixed format.

```
DATA LIST FIXED FILE="auto3.txt"
 / make 1-13 (A) mpg 15-16 weight 18-21 price 23-26
LIST
```

The output of this is shown below. The results suggest that the data were read successfully.

```
MAKE MPG WEIGHT PRICE
AMC Concord 22 2930 4099
AMC Pacer 17 3350 4749
AMC Spirit 22 2640 3799
Buick Century 20 3250 4816
Buick Electra 15 4080 7827
```

Note that the output from **LIST** displayed all of the numeric variables with no decimal places. This is different from the prior examples (see the first example on this page). When you read fixed format data in the manner that we did, SPSS associates display formats with the variables that display the data with 0 decimal places. If any of these variables had contained data after a decimal point (e.g., if **WEIGHT** was 22.25), SPSS would not display the part after the decimal point. This might fool you into thinking that there was a problem with reading in the data, when actually there is a problem in displaying the data. If we assume that **MPG** had two decimal places, we could tell SPSS to display **MPG** with two decimal places with the **FORMATS** command.

FORMATS MPG (F5.2)

If you read fixed column data with variables with data after the decimal, be sure to use the **FORMATS** command (or an equivalent) to make sure that the places after the decimal point will be displayed.

6. Tab Delimited Files and Excel Files

SPSS supports reading a variety of file formats using the **GET TRANSLATE** command, including tab delimited files, Excel files, Lotus 1-2-3 files and dBase files.

Consider the tab delimited file below. The first line of the file contains the variable names, and the following lines contain the data for the variables. This is a common file format that would be written from a spreadsheet.

MAKE	MPG	WEIGHT	PRICE
AMC Concord	22	2930	4099
AMC Pacer	17	3350	4749
AMC Spirit	22	2640	3799
Buick Century	20	3250	4816
Buick Electra	15	4080	7827

NOTES

The example below shows how to read this kind of file. The **TYPE=TAB** subcommand tells SPSS that the file is a tab delimited file, and the **FIELDNAMES** subcommand tells SPSS to expect the names of the variables in the first line of the data. We will omit the output of the commands below.

```
GET TRANSLATE FILE="auto4.txt"
 /TYPE=TAB
 /FIELDNAMES
LIST
```

You can read a tab delimited file that does not have the variable names on the first line. In that case, you would omit the **FIELDNAMES** subcommand, and you would probably want to use the **RENAME VARIABLES** command to give your variables more meaningful names (SPSS will name the variables **var1**, **var2**, **var3**, etc.). The example below shows the use of the **RENAME VARIABLES** command to rename the variables from **var1** to **var4** to **MAKE MPG WEIGHT** and **PRICE**. We omit the output of the commands below.

```
GET TRANSLATE FILE="auto5.txt"
 /TYPE=TAB
RENAME VARIABLES (var1 to var4 = MAKE MPG WEIGHT PRICE)
LIST
```

Reading an Excel file is much like reading a tab delimited file, except that you would specify the **TYPE=XLS** subcommand to indicate that you are reading an Excel file. If your file has the names of the variables on the first line, then include the **FIELDNAMES** subcommand. If your file does not have the names of the variables, the variables will be named **var1**, **var2**, **var3**, etc., and you probably will want to use the **RENAME VARIABLES** command to assign meaningful names to your variables.

```
GET TRANSLATE FILE="auto.xls"
 /TYPE=XLS
 /FIELDNAMES
LIST
```

7. Reading SAS Data Files into SPSS

SAS data files (.sas7bdat and .sd7) and SAS export files (.xpt) can be read directly into SPSS using the **GET SAS** command.

```
GET SAS DATA = "d:hsb2.sas7bdat"
EXE
```

8. Reading Stata Data Files into SPSS

The **GET STATA** command was introduced in version 14 of SPSS. It allows data sets created in versions 4-8 of Stata to be read directly into SPSS.

```
GET STATA DATA = "d:hsb2.dta"
EXE
```

NOTES

If you have a Stata data set that was created with a later version of Stata, you may need to use a program such as Stat/Transfer to convert your Stata data file into an SPSS data file, or save the Stata data file as a .csv file, and then read the .csv file into SPSS. This is often not a desirable option, as variable labels and value labels will be lost.

9. Precautions While Working with SPSS

- If you read fixed column data that have data after the decimal point, be sure to use the **formats** command to tell SPSS to display the data after the decimal point.
- If you have a string (alphanumeric) variable in the middle of your data file, SPSS may assume that all the preceding variables are also string variables. In the example below, SPSS will assume that all of the variables from **a** to **f** are string variables.

```
DATA LIST LIST
 / a b c d e f (A15) g h i j
```

to indicate to SPSS that just variable **f** is a string variable, use

```
DATA LIST LIST
 / a b c d e * f (A15) g h i j
```

which tells SPSS that to treat **a** to **e** as numeric**.**

- In trying to read an Excel spreadsheet, you may need to save the file as an earlier version of Excel to be able to read it into SPSS. If you cannot do this, you may want to use a program like Stat/Transfer to convert the data.

Check Your Progress

7. State all of the characteristics of variable to view spreadsheet.
8. How is storing and retrieving of data carried out?
9. What are sets of method used to set a limit for a format of a data file to get altered/changed?

10.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. SPSS is abbreviated term for Statistical Package for the Social Sciences and is used for data management and analysis. This program is used on computers for statistical analysis in social science by government, market researchers, education researchers, health researchers and survey companies. The statistical package SPSS is used to perform quantitative research in social science because it is easy to use.

2. The reasons for why everyone should use the syntax function are:
   - It is a way of documenting and archiving everything you have done with the data material.
   - It is easy to repeat parts or all of the analysis.
   - Other people involved in the data material can easily understand what you have done and how you have done it.
   - It saves an enormous amount of time
3. There are two ways to start the SPSS program:
   - One is to simply click on the SPSS icon shown in red letters on your desktop. If you cannot find the icon, you can click Start on the bottom of your screen, then Program Files, and then SPSS.
   - Second, if you are not sure whether the computer you are using has SPSS, click Start, then Find, then Files or Folders, then type "SPSS." When the SPSS window launches, a dialogue box will pop up.
4. SPSS Statistics provides the following features:
   - Data Editor: The Data Editor is similar to multipurpose spreadsheet system and is used to define, enter, edit and display data.
   - Viewer: The Viewer helps to browse the results, show and hide selective outputs, modify the display order results, shift presentation quality tables and charts to and from other applications.
   - Multidimensional Pivot Tables: The multidimensional pivot tables display the output results in the form which look alive. Users can explore tables by rearranging rows, columns and layers. It is also easy to compare the groups. It is done by splitting the table so that only one group is displayed at a time.
   - High Resolution Graphics: High resolution, full color pie charts, bar charts, histograms, scatter plots, 3D graphics, etc., are built-in standard features.
   - Database Access: The user can directly recover information from databases by using the Database Wizard omitting the complex SQL queries.
   - Data Transformations: Transformation features help to find the data organized for analysis. You can also subset data and files to combine categories, add, aggregate, merge, split, transpose and much more.
   - Online Help: A comprehensive abstract of context sensitive Help topics are available in dialog boxes to guide the users while performing specific tasks, pop-up definitions in pivot table results, explaining statistical terms. The Statistics Coach helps the users to find the required procedures whereas Case Studies provide hands-on examples for using statistical procedures and to interpret the results.

NOTES

- Command Language: Most of the tasks in SPSS Statistics are completed with the help of simple point-and-click actions. It also provides a powerful command language which permits the user to save and automate various common tasks. The command language also provides several functionalities which are not available in the menus and dialog boxes. Complete command syntax documentation is incorporated into the overall Help system.

5. The following are the different types of windows in SPSS Statistics:
   - Data Editor: The Data Editor displays the contents of the data file. You can create new data files or modify existing data files using the Data Editor. When you open more than one data file then there is a separate Data Editor window for each opened data file.
   - Viewer: The Viewer displays all statistical results, tables and charts. The user can edit the output and save it for later use. A Viewer window opens automatically the first time the user runs a procedure to generate output.
   - Pivot Table Editor: The Pivot Table Editor modifies the output in various ways that is displayed in pivot tables. The user can edit text, swap data in rows and columns, add color, create multidimensional tables, and hide and show selective results.
   - Chart Editor: The high resolution charts and plots can be modified in chart windows. The user can change the colors, select different font types or sizes, switch the horizontal and vertical axes, rotate 3D scatter plots and even change the chart type.
   - Text Output Editor: Text output which is not displayed in pivot tables can be modified using the Text Output Editor. The user can edit the output and modify font characteristics, such as type, style, color and size.
   - Syntax Editor: The user can paste the dialog box choices into a syntax window, where the selections appear in the form of command syntax. Now edit the command syntax to use special features that are not available through dialog boxes. The user can also save these commands in a file for use in subsequent sessions.

6. Variable View: Variable View contains descriptions of the attributes of each variable in the data file. In Variable View, the variables can be added or deleted and attributes of variables can be modified including the following attributes:

```
Rows are Variables
Columns are Variable Attributes
Variable Name
Data Type
Number of Digits or Characters
```

```
Number of Decimal Places
Descriptive Variable and Value Labels
User Defined Missing Values
Column Width
Measurement Level
```

7. There are 10 characteristics to be specified under the columns of the variable view spreadsheet:
   - Name: The chosen variable name.
   - Type: The type of data. SPSS provides a default variable type once variable values have been entered in a column of the Data View.
   - Width: The width of the actual data entries.
   - Decimals: The number of digits to the right of the decimal place to be displayed for data entries.
   - Label: A label attached to the variable name.
   - Values: Labels attached to category codes.
   - Missing: Missing value codes. SPSS recognizes the period symbol as indicating a missing value.
   - Columns: Width of the variable column in the Data View.
   - Align: Alignment of variable entries.
   - Measure: Measurement scale of the variable.
8. Storing and retrieving data files are carried out using the drop down menu available on the menu bar once you select a File. In the Data Editor a data file can be saved by using the commands File → Save or Save As…as shown in the screen. The first time you save the file you have to use Save As command for saving the file at the specified location in the specific folder. When the file already exits, then you can use Save command to save the current changes or modifications.
9. The following set of methods set a limit for the format of a data file to be changed/altered:
   - Sort Variables….
   - Transpose
   - Restructure…
   - Merge Files
   - Aggregate…
   - Sort Cases…

   Finally, the Split File…, Select Cases… and Weight Cases… procedures allow using the data file in a specific format without changing its appearance in

NOTES

the Data View spreadsheet. These commands are frequently used in data analysis. The Weight Cases… command (or from the toolbar) is typically used in connection with categorical data — it internally replicates rows according to the values of a Frequency Variable. It is useful when data is provided in the form of a cross tabulation.

10.9 SUMMARY

- SPSS is abbreviated term for Statistical Package for the Social Sciences and is used for data management and analysis.
- The statistical package SPSS is used to perform quantitative research in social science because it is easy to use. The SPSS Data Editor is very valuable and is specifically designed for performing statistical tests, such as correlation, regression, *t*-test, hypotheses, Chi-square and ANalysis Of VAriance or ANOVA. It also helps a researcher to make useful data entries, find frequency counts, sort and rearrange data, etc.
- SPSS can read and write data from ASCII (American Standard Code for Information Interchange) text files including hierarchical files, other statistics packages, spreadsheets and databases. SPSS can also be used to read and write to external relational database tables using ODBC (Open Database Connectivity) and SQL (Sequential Query Language).
- The 'Data View' exhibits a view in the form of a spreadsheet as the cases (rows) and variables (columns). Only two data types can be defined in SPSS Statistics, i.e., the numeric data type and the text or 'String' data type. All data processing processes appears in sequence case-by-case through the file.
- The 'Variable View' exhibits the metadata dictionary in which each row represents a variable to display the variable name, variable label, value label(s), print width, measurement type and other associated characteristics. In both views, you can manually edit the cells, define file structure and do data entry without using the command syntax for smaller datasets.
- The reasons for why everyone should use the syntax function are:
    (i) It is a way of documenting and archiving everything you have done with the data material.
    (ii) It is easy to repeat parts or all of the analysis.
    (iii) Other people involved in the data material can easily understand what you have done and how you have done it.
    (iv) It saves an enormous amount of time.
- SPSS Statistics 17.0 is a comprehensive system for analysing data based on the GUI. SPSS Statistics can acquire data from almost any form of file

and use them to create tabulated reports, charts, plots of distributions and trends, descriptive statistics and complex statistical analyses. SPSS Statistics Base provides examples in the Help system which is automatically installed with the software.

- Features, such as auto completion, color coding, bookmarks and breakpoints. Auto completion feature provides the user a list of valid command names, subcommands and keywords hence the user spends less time referring to syntax charts. Color coding feature permits the user to spot unrecognized terms and some common syntactical errors quickly.
- The Data Editor consists of two windows. By default the Data View, which allows the data to be entered and viewed. The other window is the Variable View, which allows the types of variables to be specified and viewed. The user can toggle between the windows by clicking on the appropriate tabs on the bottom left of the screen. Data values can be entered in the Data View spreadsheet. For most analysis SPSS assumes that rows represent cases and columns variables.
- The Variable View spreadsheet serves to define the variables. Each variable definition occupies a row of this spreadsheet. As soon as data is entered under a column in the Data View, the default name of the column occupies a row in the Variable View.
- Storing and retrieving data files are carried out using the drop down menu available on the menu bar once you select a File. In the Data Editor a data file can be saved by using the commands File ? Save or Save As…as shown in the screen. The first time you save the file you have to use Save As command for saving the file at the specified location in the specific folder. When the file already exits, then you can use Save command to save the current changes or modifications.
- The data file as displayed in the Data View spreadsheet is not always organized in the proper format for a specific use. The Data drop down menu provides methods for reorganizing the structure of a data file. The first four command options from the Data drop down menu are concerned with editing or moving within the Data View spreadsheet. Date formats can be defined or variables or cases inserted.
- The data list command indicates that we are going to read a data file. The keyword list (the second one) indicates that there is one observation per line of data and that each variable is separated by spaces and/or commas. Following the / is the list of the variable names. Since make is a string (alphanumeric) variable, we use the (A15) to tell SPSS that make is a string variable (with a length of 15). The data appear between the begin data and end data commands, and the list command is used to see if the data were read correctly.

10.10 KEY WORDS

NOTES

- **SPSS:** It is abbreviated term for Statistical Package for the Social Sciences and is used for data management and analysis.
- **Data editor:** The SPSS data editor is similar to multipurpose spreadsheet system and is used for defining, entering, editing and displaying data.
- **Viewer:** It helps to browse the results, show and hide selective outputs, modify the display order results, shift presentation quality tables and charts to and from other applications.
- **Data view:** This view displays the actual data values or defined value labels.
- **Multidimensional Pivot tables:** The multidimensional pivot tables display the output results in the form which look alive. Users can explore tables by rearranging rows, columns and layers. It is also easy to compare the groups. It is done by splitting the table so that only one group is displayed at a time.
- **High resolution graphics:** High resolution, full color pie charts, bar charts, histograms, scatter plots, 3D graphics, etc., are built-in standard features.

10.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What do you mean by SPSS?
2. What are the basic steps to login into SPSS window?
3. State the features of SPSS.
4. Explain about the Variable View Spreadsheet.
5. How do you retrieve the data files in SPSS?
6. Define data editors.
7. What are the latest features added in SPSS Statistics 17.0?

Long-Answer Questions

1. Explain the significance of SPSS Statistics 17.0 software.
2. Briefly describe about the process of working with SPSS giving appropriate examples.
3. Explain the basic structure of an SPSS Data File.
4. Discuss the basic steps of data analysis in SPSS.

5. Describe briefly about the importance of data editor in SPSS.
6. Explain the significance of data view spreadsheet in the data edition in SPSS.
7. Discuss the steps for reading Spreadsheet Data, reading a Database, and reading Text Data.

NOTES

10.12 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 11 DATA DEFINITION

NOTES

Structure

11.0 INTRODUCTION

SPSS Statistics is a software package used for interactive, or batched, statistical analysis. Long produced by SPSS Inc., it was acquired by IBM in 2009. Current versions (post 2015) have the brand name: IBM SPSS Statistics.

The software name originally stood for 'Statistical Package for the Social Science (SPSS)', reflecting the original market, then later changed to Statistical Product and Service Solutions.

Nowadays, the SPSS is a widely used program for statistical analysis in social science. It is also used by market researchers, health researchers, survey companies, government, education researchers, marketing organizations, data miners, and others. The original SPSS manual (Nie, Bent & Hull, 1970) has been described as one of 'Sociology's Most Influential Books' for allowing ordinary researchers to do their own statistical analysis. In addition to statistical analysis, data management (case selection, file reshaping, creating derived data) and data

NOTES

documentation (a Metadata Dictionary is stored in the Datafile) are features of the base software.

The many features of SPSS Statistics are accessible via pull-down menus or can be programmed with a proprietary 4GL command syntax language. Command syntax programming has the benefits of reproducible output, simplifying repetitive tasks, and handling complex data manipulations and analysis. Additionally, some complex applications can only be programmed in syntax and are not accessible through the menu structure. The pull-down menu interface also generates command syntax: this can be displayed in the output, although the default settings have to be changed to make the syntax visible to the user. They can also be pasted into a syntax file using the 'Paste' button present in each menu.

SPSS Statistics places constraints on internal file structure, data types, data processing, and matching files, which together considerably simplify programming. SPSS datasets have a two-dimensional table structure, where the rows typically represent cases (such as, individuals or households) and the columns represent measurements (such as, age, sex, or household income). Only two data types are defined: 'Numeric' and 'Text' or 'String'. All data processing occurs sequentially case-by-case through the file (dataset).

The Graphical User Interface (GUI) has two views which can be toggled by clicking on one of the two tabs in the bottom left of the SPSS Statistics window. The 'Data View' shows a spreadsheet view of the cases (rows) and variables (columns). Unlike spreadsheets, the data cells can only contain numbers or text, and formulas cannot be stored in these cells. The 'Variable View' displays the Metadata Dictionary where each row represents a variable and shows the variable name, variable label, value label(s), print width, measurement type, and a variety of other characteristics. Cells in both views can be manually edited, defining the file structure and allowing data entry without using command syntax.

In this unit, you will study about the Data Definition, Preparation of Code Book, Data List, Begin Data, End Data, Defining Variable Type, Variable Tables, Value Labels and Missing Labels.

11.1 OBJECTIVES

After going through this unit, you will be able to;

- Understand the significance of Data Definition in SPSS
- Explain and use the commands Data List, Begin Data and End Data
- Prepare Code Book in SPSS Software
- Define Variable Type, Variable Tables, Value Labels and Missing Labels in SPSS Software

11.2 INTRODUCTION TO DATA DEFINITION AND DATA ANALYSIS

NOTES

SPSS looks like Microsoft Excel, due to the cells and columns that it has. The data can be entered same as it is entered into the MS Excel spreadsheet. But with SPSS, the user has to define the variables very carefully, or else there will be error in the output results, in the graphs and in the statistical tests.

In SPSS, the Data Definition commands helps to:

- Bring raw data into SPSS, either from another file, or by typing it in the SPSS Data File.
- Enter descriptive information about the data.

Consequently, in SPSS the Data Definition includes the commands DATA LIST, BEGIN DATA and END DATA.

DATA LIST: It defines raw data that the user wants to analyse.

DATA LIST [FILE=filename]

[RECORDS={1}] /{1} varlist {col. location}

[varlist {col. location}...] [/{2 } ...] [/ ...]

- Small amounts of raw data can be entered directly during the period if the data lines are inserted between BEGIN DATA and END DATA commands.
- Large amounts of raw data are usually read from a separate ASCII file that is specified in the DATA LIST subcommand [FILE=filename].

Execute DATA LIST before selecting other commands so that the specified variable names will be available on the Variables menu.

BEGIN DATA: The signals define the beginning of a stream of raw data, as opposite to SPSS commands.

- The BEGIN DATA command begins in column 1 and should not be followed by a period.
- It is inserted alone on the line after DATA LIST and directly before the first line of raw data.
- Type the raw data for each variable in the columns for that variable as specified on the DATA LIST command.

END DATA: The signals refer to the end of a raw data stream.

The END DATA command begins in column 1 and must be followed by a period. It is inserted alone on the line after the last line of raw data.

BEGIN DATA
1234567890
9876543210
[etc.]
END DATA.

NOTES

11.3 CODE BOOK

A *Codebook* is a document containing information about each of the variables in your dataset, such as:

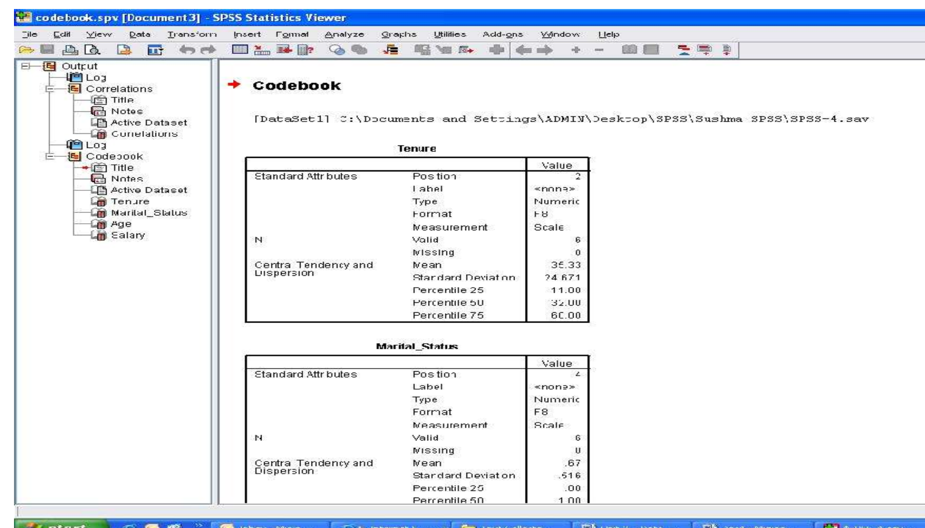
- The name assigned to the variable
- What the variable represents (i.e., its label)
- How the variable was measured (e.g. nominal, ordinal, scale)
- How the variable was actually recorded in the raw data (i.e. numeric, string; how many characters wide it is; how many decimal places it has)
- For Scale Variables: The variable's units of measurement
- For Categorical Variables: If coded numerically, the numeric codes and what they represent

11.3.1 Preparation of Code Book

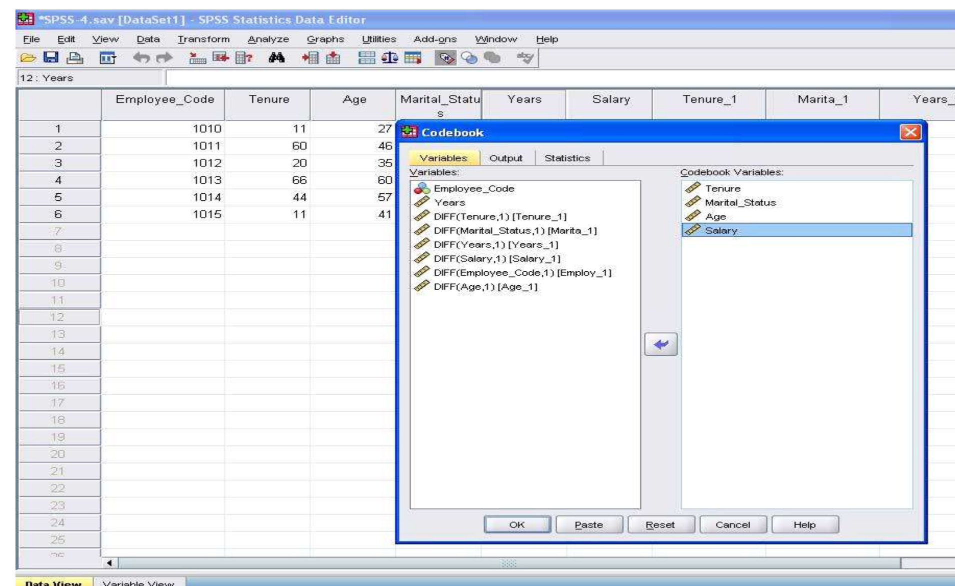
Codebook reports the dictionary information, such as variable names, variable labels, value labels, missing values and summary statistics for all or specified variables and multiple response sets in the active dataset. For nominal and ordinal variables and multiple response sets, summary statistics include counts and percent. For scale variables, summary statistics include mean, standard deviation and quartiles. Codebook ignores split file status. This includes split file groups created for multiple imputation of missing values available in the missing values Add-ons option. Select the SPSS table and then from the main menu select

Analyze → Reports → Codebook. The following screen will appear. Make the selections as per your requirements.

NOTES



When you define properties then the following screen will appear in output:



Codebooks can also contain documentation about when and how the data was created. A good Codebook allows you to communicate your research data to others clearly and succinctly, and ensures that the data is understood and interpreted properly. Different Codebooks are created manually; however, in SPSS, it's possible to generate a Codebook from an existing SPSS data file.

Check Your Progress

1. Define Codebook.
2. Write the command for preparation of Codebook.

11.4 DATA LIST

NOTES

11.4.1 Beginning of Data in a Data View

In Data View, you can enter data directly in the Data Editor in any sequence. Data can be entered either by case or by variable specifically for selected areas or for individual cells. To enter data follow the given steps:

- Highlight the active cell. The variable name and row number of the active cell are displayed in the top left corner of the Data Editor.
- Select a cell and enter a data value. The value is displayed in the cell editor at the top of the Data Editor. Data values are not recorded until you press Enter or select another cell.
- First define the variable type and then enter anything other than simple numeric data.

When a value is entered in a blank column then a new variable is automatically created. The Data Editor assigns the variable name.

To Enter Numeric Data

The numeric data is entered as follows:

- In Data View select a cell.
- Now enter the data value which will be displayed in the cell editor at the top of the Data Editor.
- Press Enter or select another cell to record the value.

To Enter Non-Numeric Data

To enter non-numeric data double click a variable name at the top of the column in Data View or click the Variable View tab as follows:

- Click the button in the Type cell for the variable.
- Select the data type in the Variable Type dialog box.
- Click OK.
- Double click the row number or click the Data View tab.
- Enter the data in the column for the recently defined variable.

Data Value Restrictions in the Data Editor

The defined variable type and width determine the type of value that can be entered in the cell in Data View.

- When you type a character which is not acceptable by the defined variable type, then the character is not entered.
- For string variables, characters beyond the defined width are not permitted.

NOTES

- For numeric variables, integer values that exceed the defined width can be entered, but the Data Editor displays either scientific notation or a portion of the value followed by an ellipsis (...) to indicate that the value is wider than the defined width. To display the value in the cell, change the defined width of the variable. Altering the column width does not have an effect on the variable width.

Editing Data and Display Log

There are several ways to edit data in SPSS. Initially, the **Data Window** rows, columns and cells can be manipulated using the '**Cut'** and **'Paste**' facilities of the **Edit** menu. Entire rows or columns can be deleted or copied or moved. Using the **Data** pull down menu you can perform the following actions:

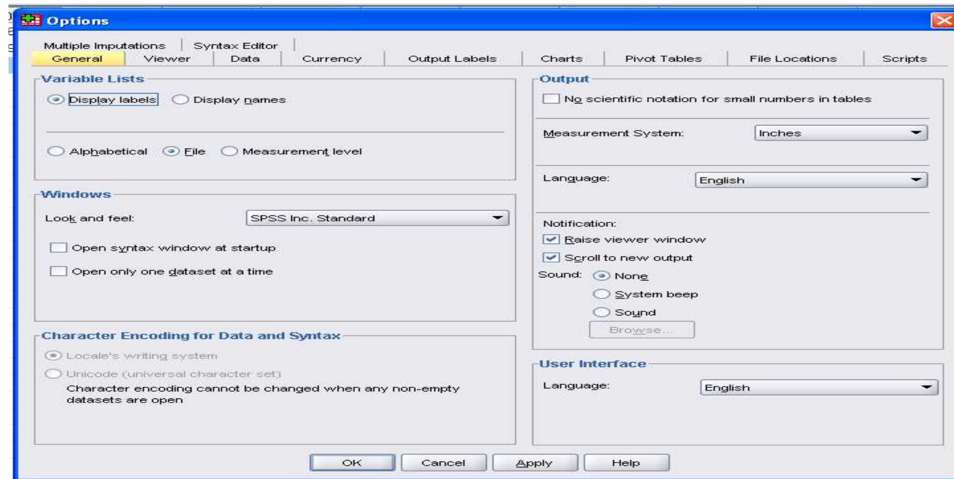
- You can select the cases by using '**Selected'** option. Selection is either temporary or permanent. Temporary selection is done using '**Filtered**' option which means that the complete dataset remains in memory even though a specific subset is being analysed whereas permanent selection is done using '**Deleted**' option which means that the complete dataset is deleted and not available for future use.
- You can sort the dataset by using '**Sorted'** option. Sorting is done according to the values of one or more variables.
- You can merge files by using the '**Merged'** option. Merging is done in two different ways, either by adding new cases (rows) to an existing file using the **'Add Cases'** option or by adding variables (columns) to the existing file using the **'Add Variables'** option.
- You can weigh the cases using the **'Weight Cases'** option. The cases are weighted in a different way during an analysis.

You can **Transform** or combine variables to create new variables. The **Compute** and **Recode** commands that are available in the **Transform** pull down menu help you to edit the data files**.** Thus, with the Data Editor you can modify data values in Data View using the following options:

- Change Data Values
- Cut, Copy and Paste Data Values
- Add and Delete Cases
- Add and Delete Variables
- Change the Order of Variables

There are numerous options available in SPSS that permit the user to modify the window displays and also the output format. These options can be accessed by selecting from the main menu **Edit → Options**. Check that **variable lists** show the **name** of the variable and not its **label**. You can also change the font

options and other output display characteristics. If you select **General** in the Option dialog box then you have to specify the given options as shown below:



Editing Output in SPSS

Employ the **outline** which is visible on the left side of the Output Viewer and **select** one or more objects, **hide** or **delete** as per your requirement. You can hide or reveal objects just by double clicking on the corresponding icon. The **Delete key** works on objects that have been selected. Right clicking the mouse helps in navigating the outline. You can also drag the objects to new positions in the outline. To initiate editing, double click on any object in the right side of the output.

Column widths can be modified by clicking on the column separator and dragging it. You can also change or modify the text within a table by double clicking on it. To change fonts use **Formatting Toolbar**. To activate the Formatting Toolbar, from the main menu select **View→ Toolbar**. The display can be changed by selecting the items that you want to modify and selecting **Cell Properties** from the **Format** pull down menu. You can also insert new lines of plain text or new titles and headings into the output between objects using the **Insert** pull down menu.

Generally, SPSS produces discrete output objects rather than ordinary ASCII text. These output objects are immensely formatted tables and are termed as **Pivot Tables** or **Charts** which can be easily edited within SPSS.

11.4.2 Replacing or Modifying Data Values

You can replace or modify the data values using the following options.

To Delete the Old Value and Enter a New Value

To delete the old value and enter a new value, follow the given steps:

- In Data View, double click the cell. The cell value will be displayed in the cell editor.
- Edit the value directly in the cell or in the cell editor.
- Press Enter or select another cell to record the new value.

NOTES

Cutting, Copying and Pasting Data Values

You can cut or copy and paste individual cell values or groups of values in the Data Editor. To move or copy the cell values, follow the given steps:

- Move or copy a single cell value to another cell
- Move or copy a single cell value to a group of cells
- Move or copy the values for a single case (row) to multiple cases
- Move or copy the values for a single variable (column) to multiple variables
- Move or copy a group of cell values to another group of cells

Data Conversion for Pasted Values in the Data Editor

When the defined variable types of the source and target cells are not identical then the Data Editor tries to convert the value. In case the conversion is not possible, the system missing value is inserted in the target cell.

Converting Numeric or Date Formats into String: Numeric and date formats can be converted to strings when they are pasted into a string variable cell. The string value is the numeric value as displayed in the cell. Values that exceed the defined string variable width are truncated.

Converting String into Numeric or Date Format: String values which contain characters that are acceptable for the numeric or date format of the target cell can be converted to the equivalent numeric or date value. For example, a string value of 25/5/11 can be converted to a valid date if the format type of the target cell is one of the day-month-year formats, but the value will be converted to system missing if the format type of the target cell is one of the month-day-year formats.

Converting Date and Time Values into Numeric Format: Date and time values can be converted to a number of seconds if the target cell is one of the numeric formats type. Dates are stored internally as the number of seconds since October 14, 1582, converting dates to numeric values can yield some extremely large numbers. For example, the date 12/23/11 can be converted to a numeric value of 13,543,977,600 as shown below.

Converting Numeric Values into Date or Time Format: Numeric values can also be converted to date or time if the value represents a number of seconds that can produce a valid date or time. For dates, numeric values that are less than 86,400 are converted to the system missing value.

Inserting New Cases

When you enter data in a cell in a blank row then a new case is automatically created. For that case the Data Editor inserts the system missing value for all other variables. If there are blank rows between the new case and the existing cases, then the blank rows become new cases with the system missing value for all variables. New cases can also be inserted between existing cases.

Inserting New Cases between Existing Cases

To insert new cases in Data View, select any cell in the case (row) below the position where the new case is to be inserted. To insert a case, from the main menu select **Edit → Insert Cases**. You will see that a new row is inserted for the case and all variables receive the system missing value.

Inserting New Variables

Entering data in an empty column in Data View or in an empty row in Variable View automatically creates a new variable with a default variable name, such as the prefix VAR and a sequential number as shown below, and a default data format type, i.e., numeric. The Data Editor inserts the system missing value for all cases for the new variable. If there are any empty columns in Data View or empty rows in Variable View between the new variable and the existing variables, these rows or columns also become new variables with the system missing value for all cases. New variables can also be inserted between existing variables.

SPSS-3.sav [DataSet0] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

11 : VAR00004

	VAR00001	VAR00002	VAR00003	VAR00004	var
1	456.00	643.00	12/23/11	13,543,977,600	
2	768.00	987.00	.	.	
3	.	.	.	.	

Inserting New Variables between Existing Variables

To insert new variables between existing variables select any cell in the variable to the right of (Data View) or below (Variable View) the position where the new variable is to be inserted. To insert a new variable, from the main menu select **Edit→ Insert Variable.** You will see that a new variable is inserted with the system missing value for all cases.

SPSS-3.sav [DataSet0] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

11 : VAR00004

	VAR00001	VAR00002	VAR00003	VAR00004	

To Change Data Type

The data type for a variable can be changed at any time by using the Variable Type dialog box in Variable View. Using the Data Editor you can convert the existing values to the new type. The system missing value is automatically allocated if no conversion is feasible. All the conversion rules are identical for entering or pasting data values to a variable with a different format type. The modification in data format

can result in the failure of missing-value specifications or value labels. The Data Editor shows an alert box and inquires whether the changes should be accepted or canceled before you proceed further.

NOTES

Finding Cases, Variables or Imputations

The '**Go To'** dialog box finds the specified case (row) number or variable name in the Data Editor.

To Find Cases: To find case, from the main menu select **Edit → Go to Case...** Enter an integer value that represents the current row number in Data View. The current row number for a particular case can change because of sorting and other actions.

To Find Variables: To find variables, from the menu select **Edit → Go to Variable...** Enter the variable name or select the variable from the drop down list.

To Find Imputations: This attribute needs the Missing Values option. From the main menu select **Edit → Go to Imputation...** Select the imputation or original data from the drop down list. Alternatively, you can select the imputation from the drop down list in the edit bar in Data View of the Data Editor.

Finding and Replacing Data and Attribute Values

To find and/or replace data values in Data View or attribute values in Variable View, click a cell in the column you want to search. Finding and replacing values is restricted to a single column only. To find a data, from the main menu, select **Edit → Find.**

To replace a data, from the main menu, select **Edit → Replace**.

Data Editor

The Data Editor provides a convenient, spreadsheet like method for creating and editing data files. The Data Editor window opens automatically when the user starts SPSS. The Data Editor provides the following two views of the data:

- **Data View:** This view displays the actual data values or defined value labels.
- **Variable View:** This view displays variable definition information including defined variable and value labels, data type (for example, string, date or numeric), measurement level (nominal, ordinal or scale) and user defined missing values.

In both views, the user can add, change and delete information contained in the data file.

- **Data View:** Many of the features of Data View are similar to the features that are found in spreadsheet applications. The following are the significant differences:
  - o **Rows are Cases:** Each row represents a case or an observation. For example, each individual respondent to a questionnaire is a case.

NOTES

- o **Columns are Variables:** Each column represents a variable or characteristic that is being measured. For example, each item on a questionnaire is a variable.
- o **Cells Contain Values:** Each cell contains a single value of a variable for a case. Where the case and the variable intersect is called the cell. Cells contain only data values. Unlike spreadsheet programs, cells in the Data Editor cannot contain formulas.
- o **The Data File is Rectangular:** The dimensions of the data file are determined by the number of cases and variables. You can enter data in any cell. If you enter data in a cell outside the boundaries of the defined data file, the data rectangle is extended to include any rows and/or columns between that cell and the file boundaries. There are no 'empty' cells within the boundaries of the data file. For numeric variables, blank cells are converted to the system missing value. For string variables, a blank is considered a valid value.

- **Variable View:** Variable View contains descriptions of the attributes of each variable in the data file. In Variable View,
  - o Rows are Variables.
  - o Columns are Variable Attributes.

The variables can be added or deleted and attributes of variables can be modified including the following attributes:

- o Variable Name
- o Data Type
- o Number of Digits or Characters
- o Number of Decimal Places
- o Descriptive Variable and Value Labels
- o User Defined Missing Values
- o Column Width
- o Measurement Level

All of these attributes are saved when the user saves the data file. In addition to defining variable properties in Variable View, the following are two other methods for defining variable properties:

- **The Copy Data Properties Wizard:** This wizard provides the ability to use an external SPSS Statistics data file or another dataset that is available in the current session as a template for defining file and variable properties in the active dataset. The user can also use variables in the active dataset as templates for other variables in the active dataset. Copy Data Properties is available on the Data menu in the Data Editor window.

NOTES

- **Define Variable Properties:** This is also available on the Data menu in the Data Editor window. It scans the specific data and lists all unique data values for any selected variables, identifies unlabelled values and provides an auto label feature. This method is used for categorical variables that use numeric codes to represent categories, for example, 0 = Male, 1 = Female.

Check Your Progress

3. Explain the steps to enter data in the data view.
4. What are the two different types of views in the Data Editor?
5. How do we convert the Numeric or Date Formats into String?

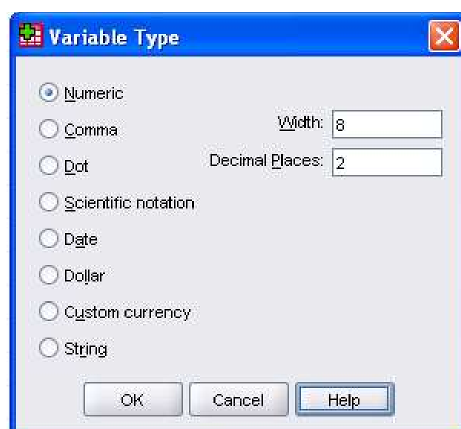
11.5 DEFINING VARIABLE TYPE

Variable Type denotes the data type for every variable. All the new variables by default are considered to be numeric. Variable Type is used to modify the data type. The Variable Type dialog box contents are dependent on the chosen data type. There are text boxes for width and number of decimals for some specific data types while for other data types you can just only select a format from a scrollable list of examples.

The available data types in SPSS Statistics are as follows:

- **Numeric:** A numeric is a variable whose values are numbers. Numeric values are displayed in standard numeric format. The Data Editor accepts numeric values in standard format or in scientific notation.
- **Comma:** A comma is a numeric variable whose values are displayed with commas delimiting every three places and displayed using the period as a decimal delimiter. The Data Editor accepts numeric values for comma variables with or without commas or in scientific notation. Remember that values cannot contain commas to the right of the decimal indicator.
- **Dot:** A dot is a numeric variable whose values are displayed with periods delimiting every three places and using the comma as a decimal delimiter. The Data Editor accepts numeric values for dot variables with or without periods or in scientific notation. Remember that values cannot contain periods to the right of the decimal indicator.
- **Scientific Notation:** A scientific notation is a numeric variable whose values are displayed using an embedded E and a signed power of 10 exponent. The Data Editor accepts numeric values for such variables with or without an exponent. The exponent can be preceded by E or D with an optional sign or by the '+' sign alone, for example, 123, 1.23E2, 1.23D2, 1.23E+2 and 1.23+2.

- **Date:** A date is also a numeric variable whose values are displayed in one of several calendar dates or clock time formats. Select the required layout or format from the list. You can enter dates using slashes, hyphens, periods, commas or blank spaces as delimiters. The century range for two-digit year values is determined using Options settings. For this, from the **Edit menu** select **Options Data Tab**.
- **Dollar:** A dollar is a numeric variable displayed with a leading dollar sign ($), commas delimiting every three places and a period as the decimal delimiter. You can enter data values with or without the leading dollar sign.
- **Custom Currency:** A custom currency is a numeric variable whose values are displayed in one of the custom currency formats that can be defined on the Currency tab of the Options dialog box. Defined custom currency characters cannot be used in data entry but are only displayed in the Data Editor.
- **String:** A string is a variable whose values are not numeric and therefore are not used in calculations. The values can contain any characters up to the defined length. Uppercase and lowercase letters are considered discrete. This type is also known as an alphanumeric variable.



To Define Variable Type

To define variable type, follow the given steps:

- Set the Data Editor as the active window.
- In Data View, double click the variable name at the top of the column in Data View or click the Variable View tab.
- Click the button in the Type cell for the variable that is to be defined.
- Select the data type in the Variable Type dialog box.
- Click OK.

NOTES

To Display or Define Variable Attributes

To display or define variable attributes, follow the given steps:

- Make the Data Editor the active window.
- Double click a variable name at the top of the column in Data View or click the Variable View tab.
- To define new variables, enter a variable name in any blank row.
- Select the attribute(s) that you want to define or modify.

Defining Variable Properties

Define Variable Properties is uniquely designed to help you in the process of creating descriptive value labels for categorical (nominal, ordinal) variables. Variable Properties is specially defined which:

- Scans the actual data values and lists all unique data values for each selected variable.
- Identifies unlabelled values and provides an 'Auto-Label' feature.
- Provides the ability to copy defined value labels from another variable to the selected variable or from the selected variable to multiple additional variables.

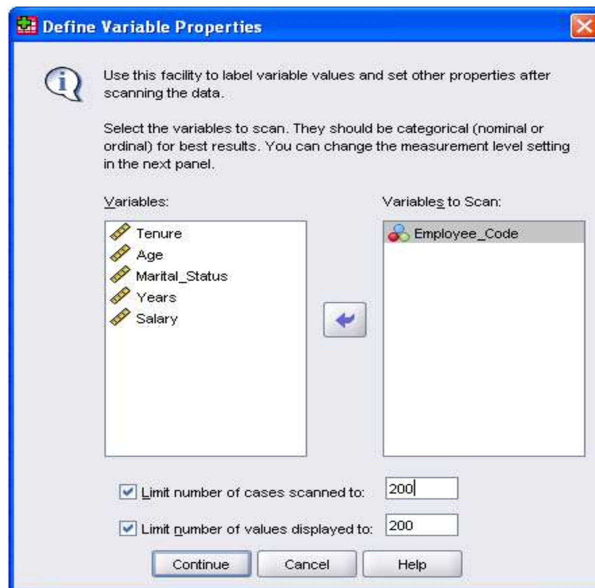
To use Define Variable Properties without first scanning cases, enter 0 for the number of cases to scan. In the first step of Define Variable Properties, perform the following actions:

- Select the variables for which you want to create value labels or define/change other variable properties, such as missing values or descriptive variable labels.
- Specify the number of cases to scan to generate the list of unique values. This is particularly useful for data files with a large number of cases for which a scan of the complete data file might take a significant amount of time.
- Specify an upper limit for the number of unique values to display. This is primarily useful to prevent listing hundreds, thousands or even millions of values for scale (continuous interval, ratio) variables.

Follow the given steps which will help you in defining variable properties:

- To define Variable Properties, from the main menu select **Data → Define Variable Properties...**
- Select the numeric or string variables for which you want to create value labels or define or change other variable properties, such as missing values or descriptive variable labels.
- Specify the number of cases to scan to generate the list of unique values. This is particularly useful for data files with a large number of cases for which a

scan of the complete data file might take a significant amount of time as shown below:



NOTES

- Specify an upper limit for the number of unique values to display. This is primarily useful to prevent listing hundreds, thousands or even millions of values for scale (continuous interval, ratio) variables.
- Click on Continue to open the main Define Variable Properties dialog box.
- Select a variable for which you want to create value labels or define or change other variable properties.
- Enter the label text for any unlabelled values that are displayed in the Value Label Grid.
- If there are values for which you want to create value labels but those values are not displayed, you can enter values in the Value column below the last scanned value.
- Repeat this process for each listed variable for which you want to create value labels.
- Click OK to apply the value labels and other variable properties. Optionally, you can also,
- Automatically determine an appropriate measurement level for a variable.
- Copy value labels and the measurement level from another variable to the selected variable.
- Copy value labels and the measurement level from the selected variable to multiple other variables.
- Auto-label values that do not have value labels using the data value as the abel.

- Designate categories as containing missing values.
- Change the variable label and/or display format.

11.5.1 Variable Names

NOTES

The following rules apply to variable names:

- Each variable name must be unique and repetition or duplication is not permitted.
- Variable names should be up to 64 bytes long and the first character have to be a letter or one of the characters @, # or $. Subsequent characters can be any arrangement of letters, numbers, non-punctuation characters and a period (.). In code page mode, sixty-four bytes characteristically defines 64 characters in single byte languages and 32 characters in double byte languages. Many string characters that only take one byte in code page mode take two or more bytes in Unicode mode. For example, é is one byte in code page format but is two bytes in Unicode format hence the word résumé is six bytes in a code page file and eight bytes in Unicode mode. Letters include any non-punctuation characters used in writing ordinary words in the languages supported in the platform's character set.
- Variable names cannot contain spaces.
- A # character in the first position of a variable name describes a scratch variable. The user can only produce scratch variables using command syntax. You cannot specify '#' as the first character of a variable in dialog boxes that create new variables.
- A $ sign in the first position specifies that the variable is a system variable. The $ sign is not allowed as the initial character of a user defined variable.
- The period, the underscore and the characters $, # and @ can be used within variable names. For example, A._$@#1 is a valid variable name.
- Variable names ending with a period should be avoided, since the period may be interpreted as a command terminator. The user is allowed to only create variables that end with a period in command syntax. The user cannot create variables that end with a period in dialog boxes that create new variables.
- Variable names that end in underscores should be avoided, since such names may automatically conflict with names of variables created using commands and procedures.
- Reserved keywords such as ALL, AND, BY, EQ, GE, GT, LE, LT, NE, NOT, OR, TO and WITH cannot be used as variable names.
- Variable names can be defined with any mixture of uppercase and lowercase characters and case is preserved for display purposes.
- When long variable names need to wrap onto multiple lines in output, lines are broken at underscores, periods and points where content changes from lower case to upper case.

Inserting or Modifying Variable Names

To insert or modify variable names, follow the given steps:

- Make the Data Editor the active window.
- Double click a variable name at the top of the column in Data View or click the Variable View tab.
- Enter the new variable name in the Name cell for the variable.

11.5.2 Variable Measurement Level

The user can also specify the level of measurement as scale (numeric data on an interval or ratio scale), ordinal or nominal as per his/her requirements. Nominal and ordinal data can be either string (alphanumeric) or numeric.

- **Nominal:** A variable can be considered as nominal when its values represent categories with no intrinsic ranking. For example, the department of the company in which an employee works.
- **Ordinal:** A variable can be considered as ordinal when its values represent categories with some intrinsic ranking. For example, levels of service satisfaction from highly dissatisfied to highly satisfy.
- **Scale:** A variable can be considered as scale when its values represent ordered categories with a meaningful metric so that distance comparisons between values are appropriate.

Note: For ordinal string variables, the alphabetic order of string values is assumed to reflect the true order of the categories. For example, for a string variable with the values of low, medium and high, the order of the categories is interpreted as high, low, medium; this is not the correct order. Generally, numeric codes are considered authentic to represent ordinal data.

You can modify the scale/nominal cut-off value for numeric variables using the Options dialog box. The Define Variable Properties dialog box, available in the Data menu, helps you to assign the correct measurement level.

New numeric variables created during a session are assigned the scale measurement level. For data read from external file formats and SPSS Statistics data files that were created prior to version 8.0, default assignment of measurement level is based on the following rules:

- Numeric variables with fewer than 24 unique values and string variables are set to nominal.
- Numeric variables with 24 or more unique values are set to scale.

Specifying Measurement Level

To specify the measurement level, follow the given steps:

- Set the Data Editor as the active window.

- Now double click on a variable name which is at the top of the column in Data View or alternatively you can click the Variable View tab.

  Choose the measurement level from the list in the Measure cell for the variable.

NOTES

11.5.3 Variable Alignment

The purpose of alignment control is to display of data values and/or value labels in Data View. The default alignment is right for numeric variables and left for string variables. This setting affects only the display in Data View.

Changing the Alignment of Values in Data View

To change the alignment of values in data view, follow the given steps:

- Set the Data Editor as the active window.
- Double click a variable name at the top of the column in Data View or alternatively click the Variable View tab.
- Select the alignment (left, right or center) from the list in the Align cell for the variable.

Getting Information about Variables in Dialog Boxes

To get information about variables in dialog boxes, follow the given steps:

- Right click a variable in the source or target variable list.
- Choose Variable Information option. It will display the information about variables.

Selecting Variables

To select a single variable, simply select it from the source variable list and then drag and drop it into the target variable list. You can also use arrow button to move variables from the source list to the target lists. If there is only one target variable list, then double click individual variables to move them from the source list to the target list. You can also select multiple variables as follows:

- To select multiple variables that are grouped together in the variable list, click the first variable and then SHIFT-CLICK the last variable in the group.
- To select multiple variables that are not grouped together in the variable list, click the first variable and then CTRL-CLICK the next variable, and so on.

11.5.4 Variable Properties

Data entered in the Data Editor in Data View or read from an external file format (such as an Excel spreadsheet or a text data file) require certain variable properties that you may find very useful including the following:

- Definition of descriptive value labels for numeric codes (for example, 0 = Male and 1 = Female).
- Identification of missing values codes (for example, 09 = Not applicable).
- Assignment of measurement level (nominal, ordinal or scale).

All of these variable properties can be assigned in Variable View in the Data Editor. Besides the following utilities can help you in the process of copying or defining properties:

- **Define Variable Properties** will help you to define descriptive value labels and missing values. This is specifically valuable for categorical data with numeric codes used for category values.
- **Copy Data Properties** provides the ability to use an existing SPSS Statistics data file as a template for file and variable properties in the current data file. This is specifically valuable if you frequently use external format data files that contain similar content, such as monthly reports in Excel format.

11.5.5 Compute Variable: Type and Label

By default, new computed variables are numeric. To compute a new string variable, you must specify the data type and width.

Label: Optional, descriptive variable label up to 255 bytes long. You can enter a label or use the first 110 characters of the compute expression as the label.

Type: Computed variables can be numeric or string (alphanumeric). String variables cannot be used in calculations.

Random Number Generators

The Random Number Generators dialog box allows you to select the random number generator and set the starting sequence value so you can reproduce a sequence of random numbers.

Count Occurrences of Values within Cases

This dialog box is used to create a variable that calculates the occurrences of the identical value(s) in a list of variables for every case. For example, a survey might contain a list of magazines with yes/no check boxes to indicate which magazines each respondent reads. You could count the number of yes responses for each respondent to create a new variable that contains the total number of magazines read.

Count Values within Cases: Values to Count

The value of the target variable on the main dialog box is incremented by '1' each time one of the selected variables matches a specification in the Values to Count List here. If a case matches several specifications for any variable, the target variable is incremented several times for that variable. Value specifications can include individual values, missing or system missing values, and ranges. Ranges include their endpoints and any user-missing values that fall within the range.

11.6 VALUE LABELS

NOTES

Value Labels: Value labels are descriptive labels associated with data values. Value labels are often used when numeric data values are used to represent non-numeric categories, for example codes of 0 and 1 for Male and Female, respectively. You can replace or merge value labels in the target variables.

- Replace deletes any defined value labels for the target variable and replaces them with the defined value labels from the source variable.
- Merge merges the defined value labels from the source variable with any existing defined value label for the target variable. If the same value has a defined value label in both the source and target variables, the value label in the target variable is unchanged.

Descriptive value labels can be assigned for each value of a variable. This process is specifically useful if your data file uses numeric codes to represent non-numeric categories, for example codes 1 and 2 for male and female, respectively. The following are the features of value label:

- Value labels are saved along with the data file. You do not need to redefine value labels each time you open a data file.
- Value labels can be up to 120 bytes.

11.6.1 How to Use Value Labels for Data Entry

When the value labels are not displayed in Data View then from the main menu select **View → Value Labels.**

- Click the cell in which the value is to be entered.
- From the drop-down list select a value label.

When you enter the value the value label is displayed in the cell. This method will only work if the value labels for the variables are defined.

To Specify Value Labels

To specify value label, follow the given steps:

- Set the Data Editor as the active window.
- In Data View, double click the variable name which is visible at the top of the column in Data View or alternatively click the Variable View tab.
- Click the button in the Values cell for the variable to be defined.
- For each value, enter the value and a label.
- Click Add to enter the value label.
- Click OK.

Additionally, you can use Define Variable Properties which will help you in the process of assigning value labels. Define Variable Properties scans the data and lists all the unique values for each variable providing a simple grid to enter value labels next to the corresponding values.

Value Label Grid

- **Label:** It displays any value labels that have already been defined. You can add or change labels in this column.
- **Value:** It displays unique values for each selected variable. This list of unique values is based on the number of scanned cases. For example, if you scanned only the first 150 cases in the data file, then the list reflects only the unique values present in those cases. If the data file has already been sorted by the variable for which you want to assign value labels, the list may display far fewer unique values than are actually present in the data.
- **Count:** It displays the number of times each value occurs in the scanned cases.
- **Missing:** It displays values defined as representing missing data. You can change the missing values designation of the category by clicking the check box. A check indicates that the category is defined as a user missing category. If a variable already has a range of values defined as user missing then you cannot add or delete missing values categories for that variable with Define Variable Properties. You can use Variable View in the Data Editor to modify the missing values categories for variables with missing values ranges.
- **Changed:** It indicates that you have added or changed a value label. To change or add a value label, simply enter the text of the value label in the Label column. If there are values for which you want to create value labels but those values are not displayed, you can enter values in the Value column below the last scanned value.

  If you specified 0 for the number of cases to scan in the initial dialog box, the Value Label grid will initially be blank except for any pre-existing value labels and/or defined missing values categories for the selected variable. Additionally, the Suggest button for the measurement level will be disabled.
- **Measurement Level:** Value labels which are primarily useful for categorical (nominal and ordinal) variables and some procedures treat categorical and scale variables differently, hence it is sometimes important to assign the correct measurement level. By default, all new numeric variables are assigned the scale measurement level. Thus, variables that are in fact categorical may initially be displayed as scale. If you are not sure of what measurement level to assign to a variable, click Suggest.

NOTES

- **Copy Properties:** You can copy value labels and other variable properties from another variable to the currently selected variable or from the currently selected variable to one or more other variables.
- **Unlabelled Values:** To create labels for unlabelled values automatically, click Automatic Labels.

11.6.2 Variable Tables

Descriptive variable labels can be assigned up to 256 characters as 128 characters in double-byte languages. Variable labels can contain spaces and reserved characters that are not permitted in variable names.

To Specify Variable Tables

To specify variable label, follow the given steps:

- Make the Data Editor the active window.
- Double click a variable name at the top of the column in Data View or click the Variable View tab.
- In the Label cell for the variable, enter the descriptive variable label.

Variable Table and Display Format

You can change the descriptive variable label and the display format. The characteristic features are as follows:

- You cannot change the variable's fundamental type, i.e., string or numeric.
- For string variables, you can change only the variable label and not the display format.
- For numeric variables, you can change the numeric type (such as numeric, date, dollar, or custom currency), width (maximum number of digits, including any decimal and/or grouping indicators) and number of decimal positions.
- For numeric date format, you can select a specific date format, (such as, dd-mm-yyyy, mm/dd/ yy and yyyyddd).
- For numeric custom format, you can select one of five custom currency formats.

An asterisk is displayed in the Value column if the specified width is less than the width of the scanned values or the displayed values for pre-existing defined value labels or missing values categories.

A period (.) is displayed if the scanned values or the displayed values for pre-existing defined value labels or missing values categories are invalid for the selected display format type. For example, an internal numeric value of less than 86,400 is invalid for a date format variable.

Custom Variable Attributes

The Attributes button in Define Variable Properties opens the Custom Variable Attributes dialog box. In addition to the standard variable attributes, such as value labels, missing values and measurement level, you can create your own custom variable attributes. Like standard variable attributes, these custom attributes are saved with SPSS Statistics data files.

Custom Variable Attributes

Name: Attribute names must follow the same rules as variable names.

Value: The value assigned to the attribute for the selected variable.

Attribute names that begin with a dollar sign are reserved and cannot be modified. You can view the contents of a reserved attribute by clicking on the button in the desired cell. The **text Array...**, displayed in a **Value Cell**, indicates that this is an attribute array, an attribute that contains multiple values. Click the button in the cell to display the list of values.

Custom Variable Attribute Arrays

The **text Array...,** displayed in a cell for a custom variable attribute in **Variable View** or in the **Custom Variable Properties** dialog box **in Define Variable Properties**, indicates that this is an **attribute array**, i.e., it is an attribute that contains multiple values. For example, you can have an attribute array that identifies the entire source variables used to compute a derived variable. Click the button in the cell to display and edit the list of values.

Copying Variable Properties

The Apply Labels and Level dialog box is displayed when you click the option **From Another Variable** or **To Other Variables** in the **Define Variable Properties** main dialog box. It displays all of the scanned variables that match the current variable's type numeric or string. For string variables, the defined width must also match. To copy variable properties follow the given steps:

- Select a single variable from which to copy value labels and other variable properties (except variable label).

  Or

- Select one or more variables to which to copy value labels and other variable properties.
- Click on Copy to copy the value labels and the measurement level.
- Existing value labels and missing value categories for target variable(s) are not replaced.
- Value labels and missing value categories for values not already defined for the target variable(s) are added to the set of value labels and missing value categories for the target variable(s).

NOTES

- The measurement level for the target variable(s) is always replaced.
- If either the source or target variable has a defined range of missing values, missing values definitions are not copied.

Custom Attributes or User-Defined Custom Variable Attributes: It helps to,

- Replace deletes any custom attributes for the target variable and replaces them with the defined attributes from the source variable.
- Merge merges the defined attributes from the source variable with any existing defined attributes for the target variable.

11.7 MISSING VALUES AND LABELS

In SPSS Statistics, Missing Values describes precise data values as user missing. The user can differentiate between data that are missing since a respondent declined to reply any precise question and data that are missing since the precise query was not appropriate to that respondent. Data values specified as user missing are considered to be flagged for particular action and can be excluded from most calculations. Missing values are values identified as representing missing data, for example, 90 for Do Not Know and 95 for Not-Applicable. Typically, these values also have defined value labels that describe what the missing value codes stand for. Any existing defined missing values for the target variable are deleted and replaced with the defined missing values from the source variable.

The following are the unique characteristics of missing values:

- User missing value specifications are saved along with the data file. You do not need to redefine user missing values each time you open the data file.
- You can enter up to three discrete (individual) missing values, a range of missing values or a range plus one discrete value.
- Ranges can be specified only for numeric variables.
- All string values, including null or blank values, are considered to be valid unless defined as missing.
- Missing values for string variables cannot exceed eight bytes. There is no limit on the defined width of the string variable, but defined missing values cannot exceed eight bytes.
- Enter a single space in one of the fields under the Discrete missing values for describing null or blank values as missing for a string variable.

11.7.1 To Define Missing Values

To define missing values, follow the given steps:

- Set the Data Editor as the active window.

- In Data View, double click the variable name at the top of the column in Data View or click the Variable View tab.
- Click the button in the Missing cell for the variable to be defined.
- Enter the values or range of values that represent missing data.
- All string values, including null or blank values, are considered to be valid unless they are explicitly defined as missing.
- Missing values for string variables cannot exceed eight bytes. There is no limit on the defined width of the string variable, but defined missing values cannot exceed eight bytes.
- To define null or blank values as missing for a string variable, enter a single space in one of the fields under the Discrete missing values selection.

Missing Values in Functions

Functions and simple arithmetic expressions consider missing values in different ways. In the expression, (var1+var2+var3)/3 the result is missing if a case has a missing value for any of the three variables.

In the expression, MEAN (var1, var2, var3) the result is missing only if the case has missing values for all three variables.

For statistical functions, you can specify the minimum number of arguments that must have non-missing values. To do so, type a period and the minimum number after the function name, such as MEAN.2 (var1, var2, var3).

11.7.2 Replace Missing Labels

Missing observations can be problematic in analysis and some time series measures cannot be computed if there are missing values in the series. Sometimes the value for a particular observation is simply not known. In addition, missing data can result from any of the following:

- Each degree of differencing reduces the length of a series by 1.
- Each degree of seasonal differencing reduces the length of a series by one season.
- If you create new series that contain forecasts beyond the end of the existing series (by clicking a Save button and making suitable choices), the original series and the generated residual series will have missing data for the new observations.
- Some transformations (for example, the log transformation) produce missing data for certain values of the original series.

Missing data at the beginning or end of a series pose no particular problem; they simply shorten the useful length of the series. Gaps in the middle of a series (embedded missing data) can be a much more serious problem. The extent of the problem depends on the analytical procedure you are using.

NOTES

The **Replace Missing Values** dialog box allows you to create new time series variables from existing ones, replacing missing values with estimates computed with one of several methods. Default new variable names are the first six characters of the existing variable used to create it, followed by an underscore and a sequential number. For example, for the variable price, the new variable name would be price_1. The new variables retain any defined value labels from the original variables.

Estimation Methods for Replacing Missing Values

Series Mean: Replaces missing values with the mean for the entire series.

Mean of Nearby Points: Missing values are replaced with the mean of valid surrounding values. The span of nearby points is the number of valid values above and below the missing value used for computing the mean.

Median of Nearby Points: Missing values are replaced with the median of valid surrounding values. The span of nearby points is the number of valid values above and below the missing value used for computing the median.

Linear Interpolation: Missing values are replaced using a linear interpolation. The last valid value before the missing value and the first valid value after the missing value are used for the interpolation. If the first or last case in the series has a missing value, the missing value is not replaced.

Linear Trend at Point: Missing values are replaced with the linear trend for that point. The existing series is regressed on an index variable scaled 1 to n. Missing values are replaced with their predicted values.

Check Your Progress

6. What do you mean by value labels?
7. Write the steps for defining variable type.
8. Explain the procedure to define the missing values in SPSS.

11.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Codebooks can also contain documentation about when and how the data was created. A good Codebook allows you to communicate your research data to others clearly and succinctly, and ensures that the data is understood and interpreted properly. Different Codebooks are created manually; however, in SPSS, it's possible to generate a Codebook from an existing SPSS data file.

2. Select the SPSS table and then from the main menu select

   Analyze → Reports → Codebook.

3. To enter data in a data view the required steps are:
   - Highlight the active cell. The variable name and row number of the active cell are displayed in the top left corner of the Data Editor.
   - Select a cell and enter a data value. The value is displayed in the cell.
   - First define the variable type and then enter anything other than simple numeric data.

4. The Data Editor provides the following two views of the data:
   - Data View: This view displays the actual data values or defined value labels.
   - Variable View: This view displays variable definition information including defined variable and value labels, data type (for example, string, date or numeric), measurement level (nominal, ordinal or scale) and user defined missing values.

   In both views, the user can add, change and delete information contained in the data file.

5. Converting Numeric or Date Formats into String: Numeric and date formats can be converted to strings when they are pasted into a string variable cell. The string value is the numeric value as displayed in the cell. Values that exceed the defined string variable width are truncated.

6. Value Labels: Value labels are descriptive labels associated with data values. Value labels are often used when numeric data values are used to represent non-numeric categories, for example codes of 0 and 1 for Male and Female, respectively. You can replace or merge value labels in the target variables.
   - Replace deletes any defined value labels for the target variable and replaces them with the defined value labels from the source variable.
   - Merge merges the defined value labels from the source variable with any existing defined value label for the target variable. If the same value has a defined value label in both the source and target variables, the value label in the target variable is unchanged.

7. To define variable type, follow the given steps:
   - Set the Data Editor as the active window.
   - In Data View, double click the variable name at the top of the column in Data View or click the Variable View tab.
   - Click the button in the Type cell for the variable that is to be defined.
   - Select the data type in the Variable Type dialog box.
   - Click OK.

NOTES

8. To define missing values, follow the given steps:
   - Set the Data Editor as the active window.
   - In Data View, double click the variable name at the top of the column in Data View or click the Variable View tab.
   - Click the button in the Missing cell for the variable to be defined.
   - Enter the values or range of values that represent missing data.
   - All string values, including null or blank values, are considered to be valid unless they are explicitly defined as missing.
   - Missing values for string variables cannot exceed eight bytes. There is no limit on the defined width of the string variable, but defined missing values cannot exceed eight bytes.
   - To define null or blank values as missing for a string variable, enter a single space in one of the fields under the Discrete missing values selection.

11.9 SUMMARY

- A *Codebook* is a document containing information about each of the variables in your dataset.
- Codebooks can also contain documentation about when and how the data was created. A good Codebook allows you to communicate your research data to others clearly and succinctly, and ensures that the data is understood and interpreted properly. Different Codebooks are created manually; however, in SPSS, it's possible to generate a Codebook from an existing SPSS data file.
- In Data View, you can enter data directly in the Data Editor in any sequence. Data can be entered either by case or by variable specifically for selected areas or for individual cells.
- To enter data follow the given steps:
  1. Highlight the active cell. The variable name and row number of the active cell are displayed in the top left corner of the Data Editor.
  2. Select a cell and enter a data value. The value is displayed in the cell editor at the top of the Data Editor. Data values are not recorded until you press Enter or select another cell.
- First define the variable type and then enter anything other than simple numeric data.
- To delete the old value and enter a new value, follow the given steps:
  1. In Data View, double click the cell. The cell value will be displayed in the cell editor.

NOTES

2. Edit the value directly in the cell or in the cell editor.
3. Press Enter or select another cell to record the new value.

- When the defined variable types of the source and target cells are not identical then the Data Editor tries to convert the value. In case the conversion is not possible, the system missing value is inserted in the target cell.
- The data type for a variable can be changed at any time by using the Variable Type dialog box in Variable View. Using the Data Editor you can convert the existing values to the new type. The system missing value is automatically allocated if no conversion is feasible. All the conversion rules are identical for entering or pasting data values to a variable with a different format type.
- The modification in data format can result in the failure of missing-value specifications or value labels. The Data Editor shows an alert box and inquires whether the changes should be accepted or cancelled before you proceed further.
- The Data Editor provides a convenient, spreadsheet like method for creating and editing data files. The Data Editor window opens automatically when the user starts SPSS. The Data Editor provides the following two views of the data:
  1. Data View: This view displays the actual data values or defined value labels.
  2. Variable View**:** This view displays variable definition information including defined variable and value labels, data type (for example, string, date or numeric), measurement level (nominal, ordinal or scale) and user defined missing values. In both views, the user can add, change and delete information contained in the data file.
- Variable Type denotes the data type for every variable. All the new variables by default are considered to be numeric. Variable Type is used to modify the data type. The Variable Type dialog box contents are dependent on the chosen data type. To insert or modify variable names, follow the given steps:
  1. Make the Data Editor the active window.
  2. Double click a variable name at the top of the column in Data View or click the Variable View tab.
  3. Enter the new variable name in the Name cell for the variable.
- Missing Values describes precise data values as user missing. The user can differentiate between data that are missing since a respondent declined to reply any precise question and data that are missing since the precise query was not appropriate to that respondent.
- Data values specified as user missing are considered to be flagged for particular action and can be excluded from most calculations.

NOTES

- Missing values are values identified as representing missing data, for example, 90 for Do Not Know and 95 for Not-Applicable. Typically, these values also have defined value labels that describe what the missing value codes stand for.
- Any existing defined missing values for the target variable are deleted and replaced with the defined missing values from the source variable.
- The following are the unique characteristics of missing values:
  1. User missing value specifications are saved along with the data file. You do not need to redefine user missing values each time you open the data file.
  2. You can enter up to three discrete (individual) missing values, a range of missing values or a range plus one discrete value.
  3. Ranges can be specified only for numeric variables.
  4. All string values, including null or blank values, are considered to be valid unless defined as missing.
  5. Missing values for string variables cannot exceed eight bytes. There is no limit on the defined width of the string variable, but defined missing values cannot exceed eight bytes.
  6. Enter a single space in one of the fields under the Discrete missing values for describing null or blank values as missing for a string variable.
- Linear Interpolation: Missing values are replaced using a linear interpolation. The last valid value before the missing value and the first valid value after the missing value are used for the interpolation. If the first or last case in the series has a missing value, the missing value is not replaced.

11.10 KEY WORDS

- **Dollar:** A dollar is a numeric variable displayed with a leading dollar sign ($), commas delimiting every three places and a period as the decimal delimiter. You can enter data values with or without the leading dollar sign.
- **Nominal:** A variable can be considered as nominal when its values represent categories with no intrinsic ranking, for example, the department of the company in which an employee works.
- **Ordinal:** A variable can be considered as ordinal when its values represent categories with some intrinsic ranking, for example, levels of service satisfaction from highly dissatisfied to highly satisfy.

- **Scale:** A variable can be considered as scale when its values represent ordered categories with a meaningful metric so that distance comparisons between values are appropriate.
- **String:** A string is a variable whose values are not numeric and therefore are not used in calculations. The values can contain any characters up to the defined length. Uppercase and lowercase letters are considered discrete. This type is also known as an alphanumeric variable.
- **Label:** Optional, descriptive variable label up to 255 bytes long. You can enter a label or use the first 110 characters of the compute expression as the label.
- **Type:** Computed variables can be numeric or string (alphanumeric). String variables cannot be used in calculations.

11.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What do you mean by Codebook?
2. How do you get information about variables in dialog boxes?
3. State the significance of defining variable type labels.
4. How to Use Value Labels for Data Entry in SPSS?
5. Write the steps for cutting, copying and pasting data values in data view.
6. Write a short note on the missing labels.
7. Explain briefly about the value label grid.

Long-Answer Questions

1. Describe the process of preparing of Codebook in SPSS.
2. Briefly describe the term value labels in SPSS.
3. Elaborate the process of finding and replacing data and attribute values in variable type.
4. Explain the estimation methods for replacing missing values in data analysis.
5. Describe the properties of variable in data analysis.
6. Explain briefly about the significance of changing alignment of variable in SPSS.

11.12 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 12 OUTPUT DISPLAY

NOTES

Structure

12.0 INTRODUCTION

SPSS utilizes multiple types of windows, or screens, in its basic operations. Each window is associated with specific tasks and types of SPSS files. The windows include the Data Editor, Output Viewer, Syntax Editor, Pivot Table Editor, Chart Editor, and Text Output Editor. Data can be in any form for further analysis. It may be in proper format or may not be. If it is not in proper format, then data must be edited before being analysed as information. This action ensures that the information provided is accurate, complete and consistent. In Data View, you can enter data directly in the Data Editor in any order by case or by variable, for selected areas or for individual cells. When a value is entered in an empty column then automatically a new variable is created and a variable name is assigned by the Data Editor. Generally SPSS produces discrete output objects rather than ordinary ASCII text. These output objects are immensely formatted tables and are termed as Pivot Tables or Charts. When the defined variable types of the source and target cells are not identical then the Data Editor tries to convert the value. In case the conversion is not possible, the system missing value is inserted in the target cell. You can change the data type for a variable at any time by using the Variable Type

dialog box in Variable View. In Data View, you can create multiple views (panes) by using the splitters that are located below the horizontal scroll bar and to the right of the vertical scroll bar. A data file can be printed in the form as it appears on the screen.

In this unit, you will study about the output display, interpretation of output, display formats, creating and editing graphs, tables and diagrams, bar, line, pie, scatter and histogram.

12.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the concept of output display
- Interpret the output data result
- Define transfer of data
- Create graphs, tables and diagrams (bar, line, pie, scatter, histogram)

12.2 OUTPUT DISPLAY

In Data View, you can enter data directly in the Data Editor in any sequence. Data can be entered either by case or by variable specifically for selected areas or for individual cells. To enter data follow the given steps:

- Highlight the active cell. The variable name and row number of the active cell are displayed in the top left corner of the Data Editor.
- Select a cell and enter a data value. The value is displayed in the cell editor at the top of the Data Editor. Data values are not recorded until you press Enter or select another cell.
- First define the variable type and then enter anything other than simple numeric data.

When a value is entered in an blank column then a new variable is automatically created. The Data Editor assigns the variable name.

To Enter Numeric Data

The numeric data is entered as follows:

- In Data View select a cell.
- Now enter the data value which will be displayed in the cell editor at the top of the Data Editor.
- Press Enter or select another cell to record the value.

NOTES

To Enter Non-Numeric Data

To enter non-numeric data double click a variable name at the top of the column in Data View or click the Variable View tab as follows:

- Click the button in the Type cell for the variable.
- Select the data type in the Variable Type dialog box.
- Click OK.
- Double click the row number or click the Data View tab.
- Enter the data in the column for the recently defined variable.

How to Use Value Labels for Data Entry

When the value labels are not displayed in Data View then from the main menu select **View** ⟶ **Value Labels.**

- Click the cell in which the value is to be entered.
- From the drop down list select a value label.

When you enter the value the value label is displayed in the cell. This method will only work if the value labels for the variables are defined.

Data Value Restrictions in the Data Editor

The defined variable type and width determine the type of value that can be entered in the cell in Data View.

- When you type a character which is not acceptable by the defined variable type, then the character is not entered.
- For string variables, characters beyond the defined width are not permitted.
- For numeric variables, integer values that exceed the defined width can be entered, but the Data Editor displays either scientific notation or a portion of the value followed by an ellipsis (...) to indicate that the value is wider than the defined width. To display the value in the cell, change the defined width of the variable. Altering the column width does not have an effect on the variable width.

Editing Data and Display Log

There are several ways to edit data in SPSS. Initially, the **Data Window** rows, columns and cells can be manipulated using the '**Cut'** and **'Paste**' facilities of the **Edit** menu. Entire rows or columns can be deleted or copied or moved. Using the **Data** pull down menu you can perform the following actions:

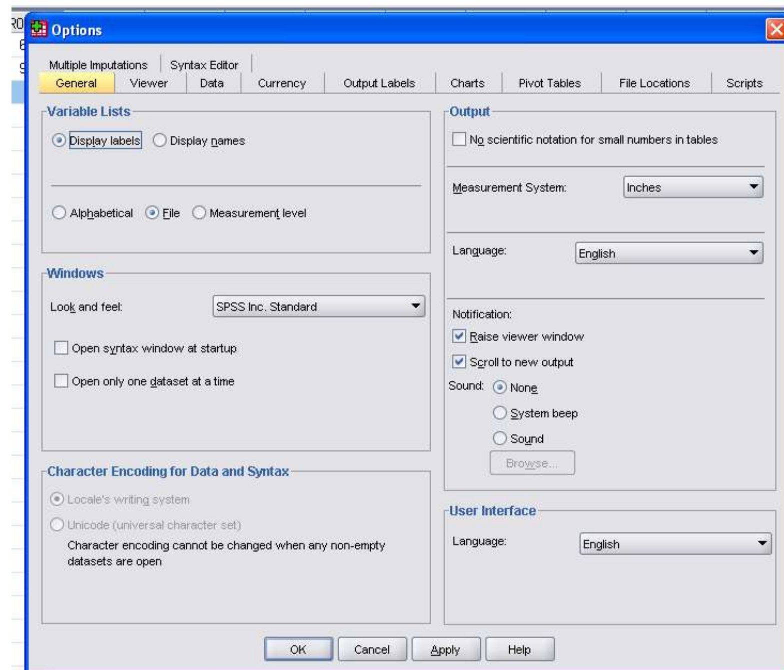
- You can select the cases by using '**Selected'** option. Selection is either temporary or permanent. Temporary selection is done using '**Filtered**' option which means that the complete dataset remains in memory even though a specific subset is being analysed whereas permanent selection is done using '**Deleted**' option which means that the complete dataset is deleted and not available for future use.

- You can sort the dataset by using '**Sorted**' option. Sorting is done according to the values of one or more variables.
- You can merge files by using the '**Merged**' option. Merging is done in two different ways, either by adding new cases (rows) to an existing file using the **'Add Cases'** option or by adding variables (columns) to the existing file using the **'Add Variables'** option.
- You can weigh the cases using the **'Weight Cases'** option. The cases are weighted in a different way during an analysis.

You can **Transform** or combine variables to create new variables. The **Compute** and **Recode** commands that are available in the **Transform** pull down menu help you to edit the data files**.** Thus with the Data Editor, you can modify data values in Data View using the following options:

- Change data values
- Cut, copy and paste data values
- Add and delete cases
- Add and delete variables
- Change the order of variables

There are numerous options available in SPSS that permit the user to modify the window displays and also the output format. These options can be accessed by selecting from the main menu **Edit** → **Options**. Check that **variable lists** show the **name** of the variable and not its **label**. You can also change the font options and other output display characteristics. If you select **General** in the Option dialog box then you have to specify the given options as shown below.



NOTES

Editing Output in SPSS

Employ the **outline** which is visible on the left side of the Output Viewer and **select** one or more objects, **hide** or **delete** as per your requirement. You can hide or reveal objects just by double clicking on the corresponding icon. The **Delete key** works on objects that have been selected. Right clicking the mouse helps in navigating the outline. You can also drag the objects to new positions in the outline. To initiate editing, double click on any object in the right side of the output.

Column widths can be modified by clicking on the column separator and dragging it. You can also change or modify the text within a table by double clicking on it. To change fonts use **Formatting Toolbar**. To activate the Formatting Toolbar, from the main menu select **View** ⟶ **Toolbar**. The display can be changed by selecting the items that you want to modify and selecting **Cell Properties** from the **Format** pull down menu. You can also insert new lines of plain text or new titles and headings into the output between objects using the **Insert** pull down menu.

Generally, SPSS produces discrete output objects rather than ordinary ASCII text. These output objects are immensely formatted tables and are termed as **Pivot Tables** or **Charts** which can be easily edited within SPSS.

12.3 REPLACING OR MODIFYING DATA VALUES

You can replace or modify the data values using the following options.

To Delete the Old Value and Enter a New Value

To delete the old value and enter a new value, follow the given steps:

- In Data View, double click the cell. The cell value will be displayed in the cell editor.
- Edit the value directly in the cell or in the cell editor.
- Press Enter or select another cell to record the new value.

Cutting, Copying and Pasting Data Values

You can cut or copy and paste individual cell values or groups of values in the Data Editor. To move or copy the cell values, follow the given steps:

- Move or copy a single cell value to another cell
- Move or copy a single cell value to a group of cells
- Move or copy the values for a single case (row) to multiple cases
- Move or copy the values for a single variable (column) to multiple variables
- Move or copy a group of cell values to another group of cells

NOTES

Data Conversion for Pasted Values in the Data Editor

When the defined variable types of the source and target cells are not identical then the Data Editor tries to convert the value. In case the conversion is not possible, the system missing value is inserted in the target cell.

Converting Numeric or Date Formats into String: Numeric and date formats can be converted to strings when they are pasted into a string variable cell. The string value is the numeric value as displayed in the cell. Values that exceed the defined string variable width are truncated.

Converting String into Numeric or Date Format: String values which contain characters that are acceptable for the numeric or date format of the target cell can be converted to the equivalent numeric or date value. For example, a string value of 25/5/11 can be converted to a valid date if the format type of the target cell is one of the day-month-year formats, but the value will be converted to system missing if the format type of the target cell is one of the month-day-year formats.

Converting Date and Time Values into Numeric Format: Date and time values can be converted to a number of seconds if the target cell is one of the numeric formats type. Dates are stored internally as the number of seconds since October 14, 1582, converting dates to numeric values can yield some extremely large numbers. For example, the date 12/23/11 can be converted to a numeric value of 13,543,977,600 as shown below:

SPSS-3.sav [DataSet0] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

11 : VAR00004

	VAR00001	VAR00002	VAR00003	VAR00004	var
1	456.00	643.00	12/23/11	13,543,977,600	
2	768.00	987.00	.	.	
3	.	.	.	.	

Converting Numeric Values into Date or Time Format: Numeric values can also be converted to date or time if the value represents a number of seconds that can produce a valid date or time. For dates, numeric values that are less than 86,400 are converted to the system missing value.

Inserting New Cases

When you enter data in a cell in a blank row then a new case is automatically created. For that case the Data Editor inserts the system missing value for all other variables. If there are blank rows between the new case and the existing cases, then the blank rows become new cases with the system missing value for all variables. New cases can also be inserted between existing cases.

NOTES

Inserting New Cases between Existing Cases

To insert new cases in Data View, select any cell in the case (row) below the position where the new case is to be inserted. To insert a case, from the main menu select **Edit** → **Insert Cases**. You will see that a new row is inserted for the case and all variables receive the system missing value.

Inserting New Variables

Entering data in an empty column in Data View or in an empty row in Variable View automatically creates a new variable with a default variable name, such as the prefix VAR and a sequential number as shown below, and a default data format type, i.e., numeric. The Data Editor inserts the system missing value for all cases for the new variable. If there are any empty columns in Data View or empty rows in Variable View between the new variable and the existing variables, these rows or columns also become new variables with the system missing value for all cases. New variables can also be inserted between existing variables.



Inserting New Variables between Existing Variables

To insert new variables between existing variables select any cell in the variable to the right of (Data View) or below (Variable View) the position where the new variable is to be inserted. To insert a new variable, from the main menu select **Edit** → **Insert Variable.** You will see that a new variable is inserted with the system missing value for all cases.

To Change Data Type

The data type for a variable can be changed at any time by using the Variable Type dialog box in Variable View. Using the Data Editor you can convert the existing values to the new type. The system missing value is automatically allocated if no conversion is feasible. All the conversion rules are identical for entering or pasting data values to a variable with a different format type. The modification in data format can result in the failure of missing-value specifications or value labels. The Data Editor shows an alert box and inquires whether the changes should be accepted or canceled before you proceed further.

Finding Cases, Variables or Imputations

The '**Go To**' dialog box finds the specified case (row) number or variable name in the Data Editor.

NOTES

To Find Cases: To find case, from the main menu **select Edit → Go to Case...** Enter an integer value that represents the current row number in Data View. The current row number for a particular case can change because of sorting and other actions.

To Find Variables: To find variables, from the menu select **Edit → Go to Variable...** Enter the variable name or select the variable from the drop down list.

To Find Imputations: This attribute needs the Missing Values option. From the main menu select **Edit → Go to Imputation...** Select the imputation or original data from the drop down list. Alternatively, you can select the imputation from the drop down list in the edit bar in Data View of the Data Editor.

Finding and Replacing Data and Attribute Values

To find and/or replace data values in Data View or attribute values in Variable View, click a cell in the column you want to search. Finding and replacing values is restricted to a single column only. To find a data, from the main menu, select **Edit → Find.**

To replace a data, from the main menu, select **Edit → Replace**.

Data View

You cannot explore for data in Data View because the search direction is always down. For dates and times, the formatted values as displayed in Data View are searched. For example, a date displayed as 10/22/2011 will not be found if you search for a date as 10-22-2011. For other numeric variables use **Contains, Begins with, Ends with** search formatted values. For example, using the **Begins with** option, a search value of $456 for a Dollar format variable will find both $456.00 and $456.70 but not $4,567. Among the Entire cell option, the search value can be formatted or unformatted, but only exact numeric values are matched. If value labels are displayed for the selected variable column, then the label text is searched and not the underlying data value. Remember that you cannot replace the label text.

Variable View

In Variable View, find is only available for the Name, Label, Values, Missing and Custom variable attribute columns. Replace is only available for the Label, Values and custom attribute columns. In the Values (value labels) column, the search string can match either the data value or a value label. Replacing the data value will delete any previous value label associated with that value.

Case Selection Status in the Data Editor

If you have selected a subset of cases but have not discarded unselected cases, unselected cases are marked in the Data Editor with a diagonal line (slash) through the row number.

The following display options are provided by View menu for the Data Editor:

Fonts: The Font option organizes the font characteristics of the data to be displayed.

Grid Lines: The Grid Lines option toggles the exhibit of grid lines.

Value Labels: The Value Label option toggles between the exhibit of real data values and user defined descriptive value labels. In Data View only this option is available.

Using Multiple Views

In Data View, you can create multiple views (panes) by using the splitters that are located below the horizontal scroll bar and to the right of the vertical scroll bar. You can also use the Window menu to insert and remove pane splitters. To insert splitters, in the main menu of **Data View** select **Window →Split**. Your Data View window will split up as shown below:

	Employee_Code	Tenure	Age	Employee_Code	Tenure	Age	Marital_Status	Years	Salary	var	var	var	var	var
1	1010	11	27	1010	11	27	1	40	23000					
2	1011	60	46	1011	60	46	0	53	45000					
3	1012	20	35	1012	20	35	1	34	21000					
4	1013	66	60	1013	66	60	0	28	30000					
5	1014	44	57	1014	44	57	1	1	10000					
6	1015	11	41	1015	11	41	1	0	5000					

Splitters can be inserted above and to the left of the selected cell. Specifically, you can use splitters as follows:

- If the top left cell is selected, splitters are inserted to divide the current view approximately in half, both horizontally and vertically.
- If any cell other than the top cell in the first column is selected, a horizontal pane splitter is inserted above the selected cell.
- If any cell other than the first cell in the top row is selected, a vertical pane splitter is inserted to the left of the selected cell.

Data Editor Printing

A data file can be printed in the form as it appears on the screen. The following are the printing specifications of data file in Data Editor:

NOTES

- You can print the information that is currently displayed. You can print the data in Data View where as in Variable View you can only print the data definition information. It is shown below.

SPSS-4.sav [DataSet2] - SPSS Statistics Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Employee_...	Numeric	8	0		None	None	12	Right	Nominal
2	Tenure	Numeric	8	0		None	None	8	Right	Scale
3	Age	Numeric	8	0		None	None	8	Right	Scale
4	Marital_Stat...	Numeric	8	0		None	None	8	Right	Scale
5	Years	Numeric	8	0		None	None	8	Right	Scale
6	Salary	Numeric	8	0		None	None	8	Right	Scale

Data View | Variable View

- You can print the Grid lines if they are currently displayed in the selected view.
- Value labels are printed in Data View if they are currently displayed. Otherwise, the actual data values are printed.

Use the View menu in the Data Editor window to display or hide grid lines and toggle between the display of data values and value labels.

To Print Data Editor Contents

To print the Data Editor contents, make the Data Editor the active window. Click the tab for the view that you want to print. To print a file, from the main menu select **File** ⟶ **Print...**

12.4 WORKING WITH MULTIPLE DATA SOURCES

Multiple data sources can be open at the same time to easily perform the following actions:

- Toggle back and forth between data sources.
- Evaluate and compare the contents of different data sources.

- Copy and paste data between data sources.
- Create multiple subsets of cases and/or variables for analysis.
- Merge multiple data sources from various data formats (for example, spreadsheet, database, text data) without saving each data source first.

NOTES

Basic Handling of Multiple Data Sources

By default, each data source that you open is displayed in a new Data Editor window. Any previously open data sources remain open and available for further use. When you first open a data source, it automatically becomes the active dataset. You can change the active dataset simply by clicking anywhere in the Data Editor window of the data source that you want to use or by selecting the Data Editor window for that data source from the Window menu. Only the variables in the active dataset are available for analysis. You cannot change the active dataset when any dialog box that accesses the data is open (including all dialog boxes that display variable lists). At least one Data Editor window must be open during a session. When you close the last open Data Editor window, SPSS Statistics automatically shuts down, prompting you to save changes first.

Working with Multiple Datasets in Command Syntax

If you use command syntax to open data sources (for example, GET FILE, GET DATA), you need to use the DATASET NAME command to name each dataset explicitly in order to have more than one data source open at the same time. When working with command syntax, the active dataset name is displayed on the toolbar of the syntax window. All of the following actions can change the active dataset:

- Use the DATASET ACTIVATE command.
- Click anywhere in the Data Editor window of a dataset.
- Select a dataset name from the toolbar in the syntax window.

Copying and Pasting Information between Datasets

You can copy both data and variable definition attributes from one dataset to another dataset in basically the same way that you copy and paste information within a single data file. The following actions can be performed:

- Copying and pasting selected data cells in Data View pastes only the data values, with no variable definition attributes.
- Copying and pasting an entire variable in Data View by selecting the variable name at the top of the column pastes all of the data and all of the variable definition attributes for that variable.
- Copying and pasting variable definition attributes or entire variables in Variable View pastes the selected attributes or the entire variable definition but does not paste any data values.

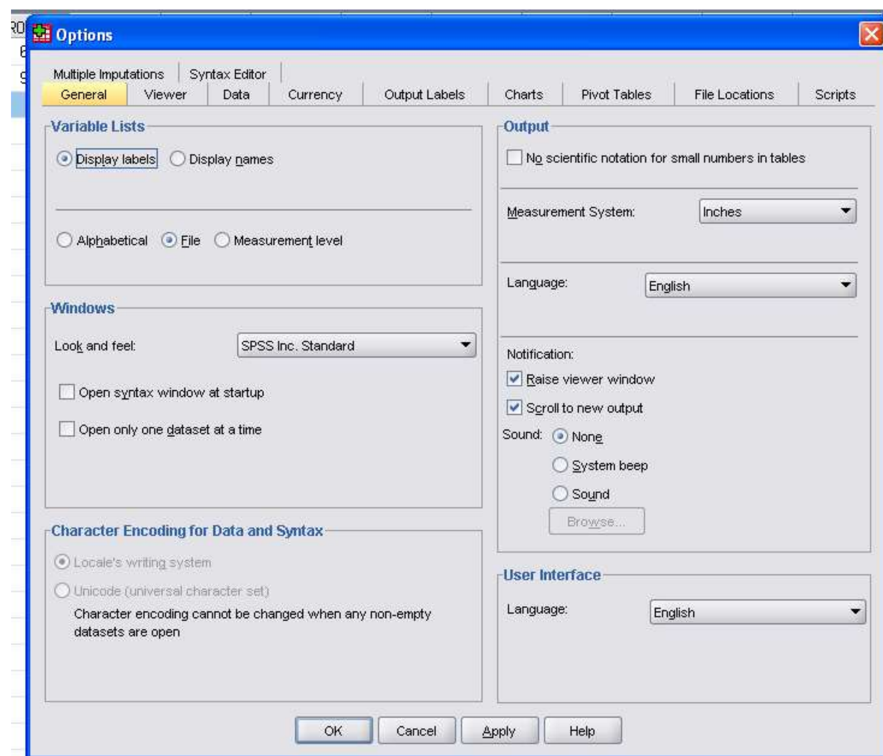
NOTES

Renaming Datasets

When you open a data source through the menus and dialog boxes, each data source is automatically assigned a dataset name of **DataSetn**, where **n** is a sequential integer value and when you open a data source using command syntax, no dataset name is assigned unless you explicitly specify one with DATASET NAME. To provide more descriptive dataset names, select for the dataset whose name you want to change. From the main menu in the Data Editor window select **File** ⟶ **Rename Dataset...** Enter a new dataset name that conforms to variable naming rules.

Suppressing Multiple Datasets

In case you want to have only one dataset available at a time and want to suppress the multiple dataset feature, from the main menu select **Edit** ⟶ **Options...** Click the General tab as shown below. Select (check) Open only one dataset at a time.



12.5 DATA PREPARATION

You can open a data file or enter data in the Data Editor and then can start creating reports, charts and analyses without any additional preliminary work. The following are some additional and useful data preparation features including the ability to:

- Assign variable properties that describe the data and determine how certain values should be treated.

- Identify cases that may contain duplicate information and exclude those cases from analyses or delete them from the data file.
- Create new variables with a few distinct categories that represent ranges of values from variables with a large number of possible values.

NOTES

Variable Properties

Data entered in the Data Editor in Data View or read from an external file format (such as an Excel spreadsheet or a text data file) require certain variable properties that you may find very useful including the following:

- Definition of descriptive value labels for numeric codes (for example, 0 = Male and 1 = Female).
- Identification of missing values codes (for example, 09 = Not applicable).
- Assignment of measurement level (nominal, ordinal or scale).

All of these variable properties can be assigned in Variable View in the Data Editor. Besides the following utilities can help you in the process of copying or defining properties:

- Define Variable Properties will help you to define descriptive value labels and missing values. This is specifically valuable for categorical data with numeric codes used for category values.
- Copy Data Properties provides the ability to use an existing SPSS Statistics data file as a template for file and variable properties in the current data file. This is specifically valuable if you frequently use external format data files that contain similar content, such as monthly reports in Excel format.

Defining Variable Properties

Define Variable Properties is uniquely designed to help you in the process of creating descriptive value labels for categorical (nominal, ordinal) variables. Variable Properties is specially defined which:

- Scans the actual data values and lists all unique data values for each selected variable.
- Identifies unlabeled values and provides an 'auto-label' feature.
- Provides the ability to copy defined value labels from another variable to the selected variable or from the selected variable to multiple additional variables.

To use Define Variable Properties without first scanning cases, enter 0 for the number of cases to scan. In the first step of Define Variable Properties, perform the following actions:

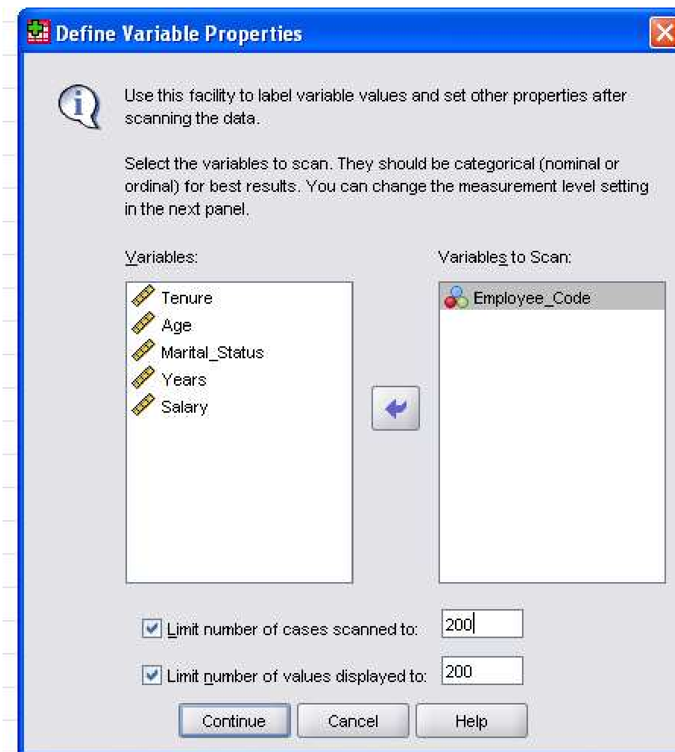
- Select the variables for which you want to create value labels or define/ change other variable properties, such as missing values or descriptive variable labels.

NOTES

- Specify the number of cases to scan to generate the list of unique values. This is particularly useful for data files with a large number of cases for which a scan of the complete data file might take a significant amount of time.
- Specify an upper limit for the number of unique values to display. This is primarily useful to prevent listing hundreds, thousands or even millions of values for scale (continuous interval, ratio) variables.

Follow the given steps which will help you in defining variable properties:

- To define Variable Properties, from the main menu select **Data** ⟶ **Define Variable Properties...**
- Select the numeric or string variables for which you want to create value labels or define or change other variable properties, such as missing values or descriptive variable labels.
- Specify the number of cases to scan to generate the list of unique values. This is particularly useful for data files with a large number of cases for which a scan of the complete data file might take a significant amount of time as shown below.



- Specify an upper limit for the number of unique values to display. This is primarily useful to prevent listing hundreds, thousands or even millions of values for scale (continuous interval, ratio) variables.
- Click on Continue to open the main Define Variable Properties dialog box.

NOTES

- Select a variable for which you want to create value labels or define or change other variable properties.
- Enter the label text for any unlabeled values that are displayed in the Value Label grid.
- If there are values for which you want to create value labels but those values are not displayed, you can enter values in the Value column below the last scanned value.
- Repeat this process for each listed variable for which you want to create value labels.
- Click OK to apply the value labels and other variable properties.

Optionally, you can also,

- Automatically determine an appropriate measurement level for a variable.
- Copy value labels and the measurement level from another variable to the selected variable.
- Copy value labels and the measurement level from the selected variable to multiple other variables.
- Auto-label values that do not have value labels using the data value as the label.
- Designate categories as containing missing values.
- Change the variable label and/or display format.

Defining Value Labels and Other Variable Properties

The Define Variable Properties main dialog box provides the following information for the scanned variables.

Scanned Variable List: For each scanned variable, a check mark in the Unlabeled (U.) column indicates that the variable contains values without assigned value labels.

Sort: To sort the variable list to display all variables with unlabeled values at the top of the list, click the 'Unlabeled' column heading under Scanned Variable List. You can also sort by variable name or measurement level by clicking the corresponding column heading under Scanned Variable List.

Value Label Grid

- **Label:** It displays any value labels that have already been defined. You can add or change labels in this column.
- **Value:** It displays unique values for each selected variable. This list of unique values is based on the number of scanned cases. For example, if you scanned only the first 150 cases in the data file, then the list reflects only the unique values present in those cases. If the data file has already been sorted by the

NOTES

variable for which you want to assign value labels, the list may display far fewer unique values than are actually present in the data.

- **Count:** It displays the number of times each value occurs in the scanned cases.
- **Missing:** It displays values defined as representing missing data. You can change the missing values designation of the category by clicking the check box. A check indicates that the category is defined as a user missing category. If a variable already has a range of values defined as user missing then you cannot add or delete missing values categories for that variable with Define Variable Properties. You can use Variable View in the Data Editor to modify the missing values categories for variables with missing values ranges.
- **Changed:** It indicates that you have added or changed a value label. To change or add a value label, simply enter the text of the value label in the Label column. If there are values for which you want to create value labels but those values are not displayed, you can enter values in the Value column below the last scanned value.

  If you specified 0 for the number of cases to scan in the initial dialog box, the Value Label grid will initially be blank except for any preexisting value labels and/or defined missing values categories for the selected variable. Additionally, the Suggest button for the measurement level will be disabled.
- **Measurement Level:** Value labels which are primarily useful for categorical (nominal and ordinal) variables and some procedures treat categorical and scale variables differently, hence it is sometimes important to assign the correct measurement level. By default, all new numeric variables are assigned the scale measurement level. Thus, variables that are in fact categorical may initially be displayed as scale. If you are not sure of what measurement level to assign to a variable, click Suggest.
- **Copy Properties:** You can copy value labels and other variable properties from another variable to the currently selected variable or from the currently selected variable to one or more other variables.
- **Unlabeled Values:** To create labels for unlabeled values automatically, click Automatic Labels.

Variable Label and Display Format

You can change the descriptive variable label and the display format. The characteristic features are as follows:

- You cannot change the variable's fundamental type, i.e., string or numeric.
- For string variables, you can change only the variable label and not the display format.

- For numeric variables, you can change the numeric type (such as numeric, date, dollar, or custom currency), width (maximum number of digits, including any decimal and/or grouping indicators) and number of decimal positions.
- For numeric date format, you can select a specific date format, (such as dd-mm-yyyy, mm/dd/yy and yyyyddd).
- For numeric custom format, you can select one of five custom currency formats.
- An asterisk is displayed in the Value column if the specified width is less than the width of the scanned values or the displayed values for preexisting defined value labels or missing values categories.
- A period (.) is displayed if the scanned values or the displayed values for preexisting defined value labels or missing values categories are invalid for the selected display format type. For example, an internal numeric value of less than 86,400 is invalid for a date format variable.

Assigning the Measurement Level

When you click Suggest for the measurement level in the Define Variable Properties main dialog box, the current variable is evaluated based on the scanned cases and defined value labels and a measurement level is suggested in the Suggest Measurement Level dialog box that opens. The Explanation area provides a brief description of the criteria used to provide the suggested measurement level. Values defined as representing missing values are not included in the evaluation for measurement level. For example, the explanation for the suggested measurement level may indicate that the suggestion is in part based on the fact that the variable contains no negative values, whereas it may in fact contain negative values but those values are already defined as missing values. Click 'Continue' to accept the suggested level of measurement or 'Cancel' to leave the measurement level unchanged.

Custom Variable Attributes

The Attributes button in Define Variable Properties opens the Custom Variable Attributes dialog box. In addition to the standard variable attributes, such as value labels, missing values and measurement level, you can create your own custom variable attributes. Like standard variable attributes, these custom attributes are saved with SPSS Statistics data files.

Custom Variable Attributes

Name: Attribute names must follow the same rules as variable names.

Value: The value assigned to the attribute for the selected variable.

Attribute names that begin with a dollar sign are reserved and cannot be modified. You can view the contents of a reserved attribute by clicking on the

button in the desired cell. The text Array..., displayed in a Value cell, indicates that this is an attribute array, an attribute that contains multiple values. Click the button in the cell to display the list of values.

NOTES

Custom Variable Attribute Arrays

The text **Array...,** displayed in a cell for a custom variable attribute in Variable View or in the Custom Variable Properties dialog box in Define Variable Properties, indicates that this is an attribute array, i.e., it is an attribute that contains multiple values. For example, you can have an attribute array that identifies the entire source variables used to compute a derived variable. Click the button in the cell to display and edit the list of values.

Copying Variable Properties

The Apply Labels and Level dialog box is displayed when you click the option **From Another Variable** or **To Other Variables** in the **Define Variable Properties** main dialog box. It displays all of the scanned variables that match the current variable's type numeric or string. For string variables, the defined width must also match. To copy variable properties follow the given steps:

- Select a single variable from which to copy value labels and other variable properties (except variable label).

  Or

- Select one or more variables to which to copy value labels and other variable properties.
- Click Copy to copy the value labels and the measurement level.
- Existing value labels and missing value categories for target variable(s) are not replaced.
- Value labels and missing value categories for values not already defined for the target variable(s) are added to the set of value labels and missing value categories for the target variable(s).
- The measurement level for the target variable(s) is always replaced.
- If either the source or target variable has a defined range of missing values, missing values definitions are not copied.

12.6 MULTIPLE RESPONSE SETS

Custom Tables and the Chart Builder support a particular type of 'variable' termed as a multiple response set. Multiple response sets are not actual variables in the standard terms. These can not be seen in the Data Editor and other methods also do not identify them. Multiple response sets employ multiple variables for recording responses to questions in case the respondent gives more than one answer. Multiple response sets are considered as similar categorical variables and you can perform

NOTES

various actions with categorical variables. You can also perform various actions with multiple response sets. In data file, the multiple response sets are created from multiple variables. A multiple response set is considered as a unique construct within a data file. You can describe and save multiple response sets in SPSS Statistics data files. Multiple response sets cannot be imported or exported from/ to other file formats. To copy multiple response sets from other SPSS Statistics data files use Copy Data Properties which can be accessed from the Data menu in the Data Editor window.

Defining Multiple Response Sets: To define multiple response sets, follow the given steps:

- From the main menu select **Data** → **Define Multiple Response Sets...**
- Select two or more variables. If your variables are coded as dichotomies, indicate which value you want to have counted.
- Enter a unique name for each multiple response set. The name can be up to 63 bytes long. A dollar sign is automatically added to the beginning of the set name.
- Enter a descriptive label for the set. This is optional.
- Click Add to add the Multiple Response Sets to the list of defined sets.

Copying Data Properties

The Copy Data Properties Wizard provides the ability to use an external SPSS Statistics data file as a template for defining file and variable properties in the active dataset. You can also use variables in the active dataset as templates for other variables in the active dataset. The following copy actions can be performed:

- Copy selected file properties from an external data file or open dataset to the active dataset. File properties include documents, file labels, multiple response sets, variable sets and weighting.
- Copy selected variable properties from an external data file or open dataset to matching variables in the active dataset. Variable properties include value labels, missing values, level of measurement, variable labels, print and write formats, alignment and column width (in the Data Editor).
- Copy selected variable properties from one variable in either an external data file open dataset or the active dataset to many variables in the active dataset.
- Create new variables in the active dataset based on selected variables in an external data file or open dataset.

The following general rules are applicable while copying data properties:

- If you use an external data file as the source data file, it must be a data file in SPSS Statistics format.
- If you use the active dataset as the source data file, it must contain at least one variable. You cannot use a completely blank active dataset as the source data file.

- Undefined (empty) properties in the source dataset do not overwrite defined properties in the active dataset.
- Variable properties are copied from the source variable only to target variables of a matching type, string (alphanumeric) or numeric (including numeric, date and currency).

Copy Data Properties replaces Apply Data Dictionary, formerly available on the File menu.

To copy Data Properties, from the main menu in the Data Editor window select **Data → Copy Data Properties...**

Select the data file with the file and/or variable properties that you want to copy. This can be a currently open dataset, an external SPSS Statistics data file or the active dataset. Follow the step-by-step instructions in the Copy Data Properties Wizard.

Selecting Source and Target Variables

This step helps you to specify the source variables containing the variable properties that you want to copy and the target variables that will receive those variable properties.

Apply properties from selected source dataset variables to matching active dataset variables. Variable properties are copied from one or more selected source variables to matching variables in the active dataset. Variables 'match' if both the variable name and type (string or numeric) are the same. For string variables, the defined length must also be the same. By default, only matching variables are displayed in the two variable lists.

Create matching variables in the active dataset if they do not already exist. This updates the source list to display all variables in the source data file. If you select source variables that do not exist in the active dataset (based on variable name), new variables will be created in the active dataset with the variable names and properties from the source data file.

If the active dataset contains no variables (a blank, new dataset), all variables in the source data file are displayed and new variables based on the selected source variables are automatically created in the active dataset.

Apply properties from a single source variable to selected active dataset variables of the same type. Variable properties from a single selected variable in the source list can be applied to one or more selected variables in the active dataset list. Only variables of the same type (numeric or string) as the selected variable in the source list are displayed in the active dataset list. For string variables, only strings of the same defined length as the source variable are displayed. This option is not available if the active dataset contains no variables. You cannot create new variables in the active dataset with this option.

Choosing Variable Properties to Copy

You can copy selected variable properties from the source variables to the target variables. Undefined (empty) properties in the source variables do not overwrite defined properties in the target variables.

Value Labels: Value labels are descriptive labels associated with data values. Value labels are often used when numeric data values are used to represent non-numeric categories, for example codes of 0 and 1 for Male and Female, respectively. You can replace or merge value labels in the target variables.

- Replace deletes any defined value labels for the target variable and replaces them with the defined value labels from the source variable.
- Merge merges the defined value labels from the source variable with any existing defined value label for the target variable. If the same value has a defined value label in both the source and target variables, the value label in the target variable is unchanged.

Custom Attributes or User-Defined Custom Variable Attributes: It helps to,

- Replace deletes any custom attributes for the target variable and replaces them with the defined attributes from the source variable.
- Merge merges the defined attributes from the source variable with any existing defined attributes for the target variable.

Missing Values: Missing values are values identified as representing missing data, for example, 90 for Do not know and 95 for Not applicable. Typically, these values also have defined value labels that describe what the missing value codes stand for. Any existing defined missing values for the target variable are deleted and replaced with the defined missing values from the source variable.

Variable Label: Descriptive variable labels can contain spaces and reserved characters not allowed in variable names. If variable properties are copied from a single source variable to multiple target variables then you have to check before selecting this option.

Measurement Level: The measurement level can be nominal, ordinal or scale. Procedures that distinguish between different measurement levels, nominal and ordinal are considered categorical.

Formats: For numeric variables, this controls numeric type (such as numeric, date or currency), width (total number of displayed characters, including leading and trailing characters and decimal indicator) and number of decimal places displayed. This option is ignored for string variables.

Alignment: This affects only alignment (left, right, center) in Data View in the Data Editor.

Data Editor Column Width. This affects only column width in Data View in the Data Editor.

NOTES

12.7 COPYING DATASET (FILE) PROPERTIES

You can apply selected, global dataset properties from the source data file to the active dataset. This is not available if the active dataset is the source data file.

Multiple Response Sets: This property applies multiple response set definitions from the source data file to the active dataset. Multiple response sets in the source data file that contain variables that do not exist in the active dataset are ignored unless those variables will be created based on specifications done in Selecting Source and Target Variables in the Copy Data Properties Wizard. Replace deletes all Multiple Response Sets in the active dataset and replaces them with the multiple response sets from the source data file. Merge adds multiple response sets from the source data file to the collection of multiple response sets in the active dataset. If a set with the same name exists in both files, the existing set in the active dataset is unchanged.

Variable Sets: Variable sets are used to control the list of variables that are displayed in dialog boxes. Variable sets are defined by selecting Define Sets from the Utilities menu. Sets in the source data file that contain variables that do not exist in the active dataset are ignored unless those variables will be created based on specifications done in Selecting Source and Target Variables in the Copy Data Properties Wizard. Replace deletes any existing variable sets in the active dataset, replacing them with variable sets from the source data file. Merge adds variable sets from the source data file to the collection of variable sets in the active dataset. If a set with the same name exists in both files, the existing set in the active dataset is unchanged.

Documents: Notes appended to the data file via the DOCUMENT command. Replace deletes any existing documents in the active dataset, replacing them with the documents from the source data file. Merge combines documents from the source and active dataset. Unique documents in the source file that do not exist in the active dataset are added to the active dataset. All documents are then sorted by date.

Custom Attributes: Custom data file attributes, typically created with the DATAFILE ATTRIBUTE command in command syntax. Replace deletes any existing custom data file attributes in the active dataset, replacing them with the data file attributes from the source data file. Merge combines data file attributes from the source and active dataset. Unique attribute names in the source file that do not exist in the active dataset are added to the active dataset. If the same attribute name exists in both data files, the named attribute in the active dataset is unchanged.

Weight Specification: Weights cases by the current weight variable in the source data file if there is a matching variable in the active dataset. This overrides any weighting currently in effect in the active dataset.

File Label: Descriptive label applied to a data file with the FILE LABEL command.

NOTES

12.8 RESULTS

The last step in the Copy Data Properties Wizard provides information on the number of variables for which variable properties will be copied from the source data file, the number of new variables that will be created, and the number of dataset (file) properties that will be copied. You can also choose to paste the generated command syntax into a syntax window and save the syntax for later use.

Identifying Duplicate Cases

'Duplicate' cases may occur in your data for many reasons including:

- Data entry errors in which the same case is accidentally entered more than once.
- Multiple cases share a common primary ID value but have different secondary ID values, such as family members who all live in the same house.
- Multiple cases represent the same case but with different values for variables other than those that identify the case, such as multiple purchases made by the same person or company for different products or at different times.

Identify Duplicate Cases allows you to define duplicate almost any way that you want and provides some control over the automatic determination of primary versus duplicate cases.

How to Identify Duplicate Cases

Define matching cases by. Cases are considered duplicates if their values match for all selected variables. If you want to identify only cases that are a 100% match in all respects, select all of the variables. Sort within matching groups by. Cases are automatically sorted by the variables that define matching cases. You can select additional sorting variables that will determine the sequential order of cases in each matching group.

- **Sequential Count of Matching Cases in Each Group:** Creates a variable with a sequential value from **1** to **n** for cases in each matching group. The sequence is based on the current order of cases in each group, which is either the original file order or the order determined by any specified sort variables.
- **Move Matching Cases to the Top:** Sorts the data file so that all groups of matching cases are at the top of the data file, making it easy to visually inspect the matching cases in the Data Editor.

NOTES

- **Display Frequencies for Created Variables:** Frequency tables containing counts for each value of the created variables. For example, for the primary indicator variable, the table would show the number of cases with a value 0 for that variable, which indicates the number of duplicates and the number of cases with a value of 1 for that variable, which indicates the number of unique and primary cases.
- **Missing Values:** For numeric variables, the system missing value is treated like any other value. Cases with the system missing value for an identifier variable are treated as having matching values for that variable. For string variables, cases with no value for an identifier variable are treated as having matching values for that variable.

12.9 DATA TRANSFORMATION

Sometimes the data that have to be analysed is not in the proper format for analysis. In such cases the data must be modified for analysis and evaluation to get the correct results. This modification is done through data transformation which includes the following steps:

Computing Variables

Use the Compute dialog box for computing values for a variable that is based on numeric transformations of other variables.

- You can compute values for numeric or string (alphanumeric) variables.
- You can create new variables or replace the values of existing variables. For new variables, you can also specify the variable type and label.
- You can compute values selectively for subsets of data based on logical conditions.
- You can use more than 70 built-in functions, including arithmetic functions, statistical functions, distribution functions and string functions.

Compute Variable: If Cases

The **If Cases** dialog box allows you to apply data transformations to selected subsets of cases using conditional expressions. A conditional expression returns a value of true, false or missing for each case.

- If the result of a conditional expression is true, the case is included in the selected subset.
- If the result of a conditional expression is false or missing, the case is not included in the selected subset.
- Most conditional expressions use one or more of the six relational operators (<, >, ≤ ≥ = and ≠) on the calculator pad.

- Conditional expressions can include variable names, constants, arithmetic operators, numeric (and other) functions, logical variables and relational operators.

Compute Variable: Type and Label

By default, new computed variables are numeric. To compute a new string variable, you must specify the data type and width.

Label: Optional, descriptive variable label up to 255 bytes long. You can enter a label or use the first 110 characters of the compute expression as the label.

Type: Computed variables can be numeric or string (alphanumeric). String variables cannot be used in calculations.

Missing Values in Functions

Functions and simple arithmetic expressions consider missing values in different ways. In the expression, (var1+var2+var3)/3 the result is missing if a case has a missing value for any of the three variables.

In the expression, MEAN(var1, var2, var3) the result is missing only if the case has missing values for all three variables.

For statistical functions, you can specify the minimum number of arguments that must have nonmissing values. To do so, type a period and the minimum number after the function name, such as MEAN.2(var1, var2, var3)

Random Number Generators

The Random Number Generators dialog box allows you to select the random number generator and set the starting sequence value so you can reproduce a sequence of random numbers.

Count Occurrences of Values within Cases

This dialog box is used to create a variable that calculates the occurrences of the identical value(s) in a list of variables for every case. For example, a survey might contain a list of magazines with yes/no check boxes to indicate which magazines each respondent reads. You could count the number of yes responses for each respondent to create a new variable that contains the total number of magazines read.

Count Values within Cases: Values to Count

The value of the target variable on the main dialog box is incremented by 1 each time one of the selected variables matches a specification in the Values to Count list here. If a case matches several specifications for any variable, the target variable is incremented several times for that variable. Value specifications can include individual values, missing or system missing values, and ranges. Ranges include their endpoints and any user-missing values that fall within the range.

NOTES

Count Occurrences: If Cases

The If Cases dialog box allows you to count occurrences of values for a selected subset of cases, using conditional expressions. A conditional expression returns a value of true, false or missing for each case.

Shift Values

Shift Values creates new variables that contain the values of existing variables from preceding or subsequent cases.

Name: Name for the new variable. This must be a name that does not already exist in the active dataset.

Get Value from Earlier Case (Lag): Get the value from a previous case in the active dataset. For example, with the default number of cases value of 1, each case for the new variable has the value of the original variable from the case that immediately precedes it.

Get Value from Following Case (Lead): Get the value from a subsequent case in the active dataset. For example, with the default number of cases value of 1, each case for the new variable has the value of the original variable from the next case.

Number of Cases to Shift: Get the value from the nth preceding or subsequent case, where n is the value specified. The value must be a non-negative integer.

- If split file processing is on, the scope of the shift is limited to each split group. A shift value cannot be obtained from a case in a preceding or subsequent split group.
- Filter status is ignored.
- The value of the result variable is set to system missing for the first or last n cases in the dataset or split group, where n is the value specified for Number of cases to shift. For example, using the Lag method with a value of 1 would set the result variable to system missing for the first case in the dataset or first case in each split group.
- User missing values are preserved.
- Dictionary information from the original variable, including defined value labels and user missing value assignments, is applied to the new variable.
- A variable label is automatically generated for the new variable that describes the shift operation that created the variable.

Recode into Same Variables

The Recode into Same Variables dialog box permits the user to again assign the values of existing variables or collapse ranges of existing values into new values. For example, you could collapse salaries into salary range categories.

You can recode numeric and string variables. If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

NOTES

Recode into Same Variables: Old and New Values

You can define values to recode in this dialog box. All value specifications must be the same data type (numeric or string) as the variables selected in the main dialog box.

Old Value: The value(s) to be recoded. You can recode single values, ranges of values, and missing values. System missing values and ranges cannot be selected for string variables because neither concept applies to string variables. Ranges include their endpoints and any user missing values that fall within the range.

New Value: The single value into which each old value or range of values is recoded. You can enter a value or assign the system missing value.

Old → New: The list of specifications that will be used to recode the variable(s). You can add, change and remove specifications from the list. The list is automatically sorted, based on the old value specification using the order single values, missing values, ranges and all other values. If you change a recode specification on the list, the procedure automatically resorts the list if necessary to maintain this order.

Recode into Same Variables: If Cases

The If Cases dialog box allows you to recode values for a selected subset of cases, using conditional expressions. A conditional expression returns a value of true, false or missing for each case. If the result of a conditional expression is true, the case is included in the selected subset.

Recode into Different Variables

The Recode into Different Variables dialog box allows you to reassign the values of existing variables or collapse ranges of existing values into new values for a new variable. For example, you could collapse salaries into a new variable containing salary range categories. You can recode numeric and string variables. You can recode numeric variables into string variables and vice versa. If you select multiple variables, they must all be the same type. You cannot recode numeric and string variables together.

Automatic Recode

The Automatic Recode dialog box allows you to convert string and numeric values into consecutive integers. When category codes are not sequential, the resulting empty cells reduce performance and increase memory requirements for many procedures. Additionally, some procedures cannot use string variables and some require consecutive integer values for factor levels.

NOTES

- The new variable(s) created by Automatic Recode retain any defined variable and value labels from the old variable. For any values without a defined value label, the original value is used as the label for the recoded value. A table displays the old and new values and value labels.
- String values are recoded in alphabetical order with uppercase letters preceding their lowercase counterparts.
- Missing values are recoded into missing values higher than any nonmissing values, with their order preserved.

  Use the same recoding scheme for all variables. This option allows you to apply a single '**autorecoding'** scheme to all the selected variables yielding a consistent coding scheme for all the new variables.

If you select this option, the following rules and limitations apply:

- All variables must be of the same type (numeric or string).
- All observed values for all selected variables are used to create a sorted order of values to recode into sequential integers.
- User missing values for the new variables are based on the first variable in the list with defined user missing values. All other values from other original variables, except for system missing, are treated as valid.

Treat blank string values as user-missing. For string variables, blank or null values are not treated as system missing. This option will autorecode blank strings into a user missing value higher than the highest nonmissing value.

Templates

You can save the autorecoding scheme in a template file and then apply it to other variables and other data files. For example, you may have a large number of alphanumeric product codes that you autorecode into integers every month, but some months new product codes are added that change the original autorecoding scheme. If you save the original scheme in a template and then apply it to the new data that contain the new set of codes, any new codes encountered in the data are autorecoded into values higher than the last value in the template, preserving the original autorecode scheme of the original product codes.

Rank Cases

The Rank Cases dialog box allows you to create new variables containing ranks, normal and Savage scores, and percentile values for numeric variables.

New variable names and descriptive variable labels are created or generated automatically, based on the original variable name and the selected measure(s). A summary table lists the original variables, the new variables and the variable labels.

The automatically generated new variable names are limited to a maximum length of 8 bytes.

Optionally, you can Rank cases in ascending or descending order. Organize rankings into subgroups by selecting one or more grouping variables for the By list. Ranks are computed within each group. Groups are defined by the combination of values of the grouping variables. For example, if you select gender and minority as grouping variables, ranks are computed for each combination of gender and minority.

Rank Cases: Types

You can select multiple ranking methods. A separate ranking variable is created for each method. Ranking methods include simple ranks, Savage scores, fractional ranks and percentiles. You can also create rankings based on proportion estimates and normal scores.

Proportion Estimates: Estimates of the cumulative proportion of the distribution corresponding to a particular rank.

Normal Scores: The z scores corresponding to the estimated cumulative proportion.

Proportion Estimation Formula: For proportion estimates and normal scores, you can select the proportion estimation formula: Blom, Tukey, Rankit or Van der Waerden.

Rank Cases: Ties

This dialog box controls the method for assigning rankings to cases with the same value on the original variable. The given table shows how the different methods assign ranks to tied values:

Value	Mean	Low	High	Sequential
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

12.10 SCATTER DIAGRAM METHOD

This method makes use of the Scatter diagram also known as Dot diagram. *Scatter diagram* is a diagram representing two series with the known variable, i.e.,

independent variable plotted on the X-axis and the variable to be estimated, i.e., dependent variable to be plotted on the Y-axis on a graph paper (refer Figure 12.1) to get the following information:

NOTES

Income X *(Hundreds of Rupees)*	*Consumption Expenditure* Y *(Hundreds of Rupees)*
41	44
65	60
50	39
57	51
96	80
94	68
110	84
30	34
79	55
65	48

The scatter diagram by itself is not sufficient for predicting values of the dependent variable. Some formal expression of the relationship between the two variables is necessary for predictive purposes. For the purpose, one may simply take a ruler and draw a straight line through the points in the scatter diagram and this way can determine the intercept and the slope of the said line and then the line can be defined as $\hat{Y} = a + bX_i$, with the help of which we can predict Y for a given value of X. But there are shortcomings in this approach. For example, if five different persons draw such a straight line in the same scatter diagram, it is possible that there may be five different estimates of a and b, specially when the dots are more dispersed in the diagram. Hence, the estimates cannot be worked out only through this approach. A more systematic and statistical method is required to estimate the constants of the predictive equation. The least squares method is used to draw the best fit line.

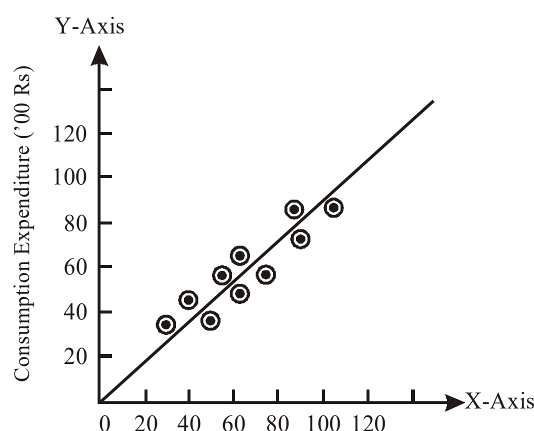
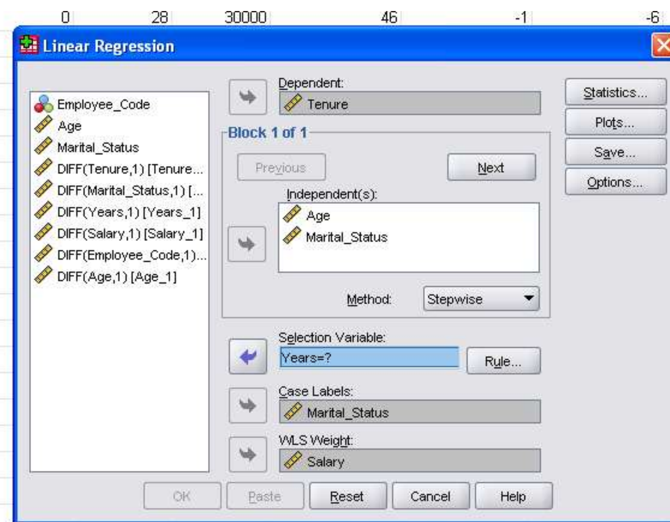


Fig. 12.1 *Scatter Diagram*

Linear Regression

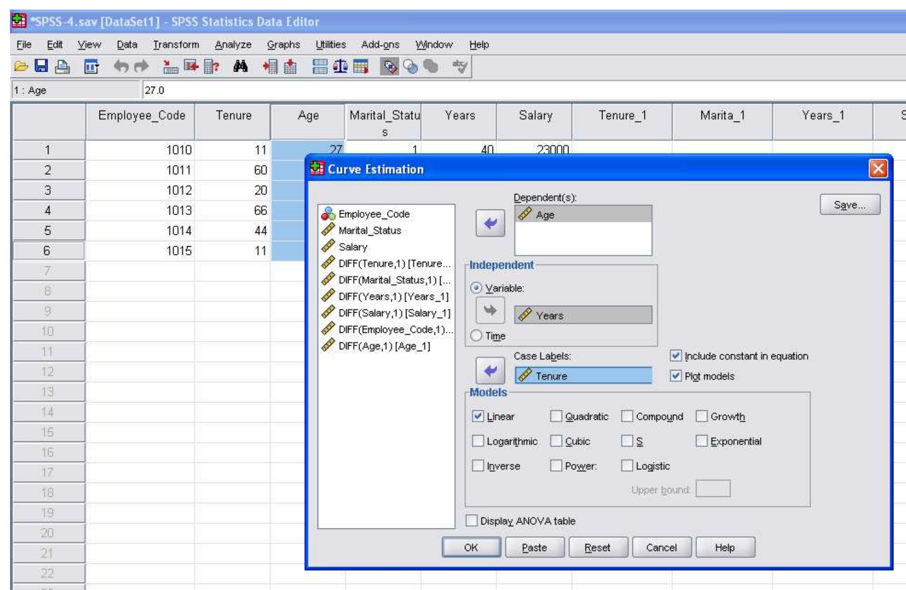
Linear Regression estimates the coefficients of the linear equation, involving one or more independent variables that best predict the value of the dependent variable. For example, you can try to predict a salesperson's total yearly sales (the dependent variable) from independent variables such as age, education, and years of experience.

NOTES

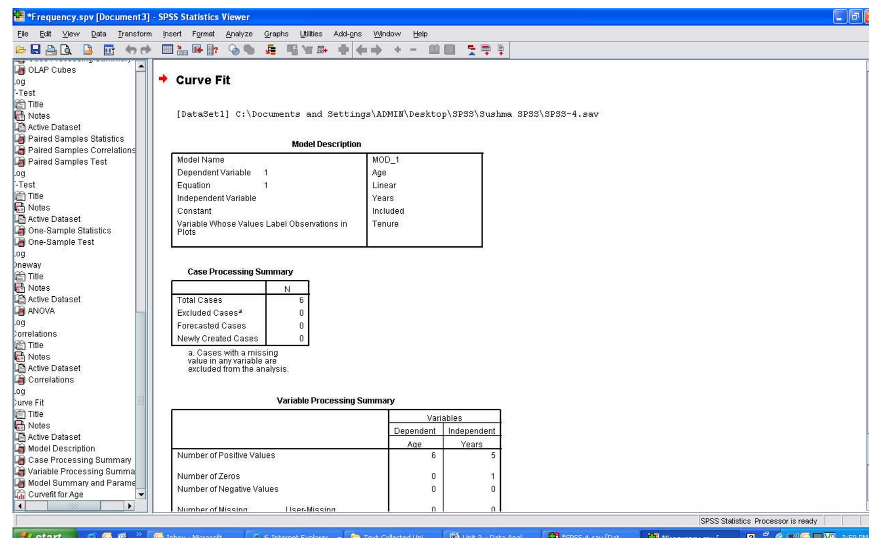


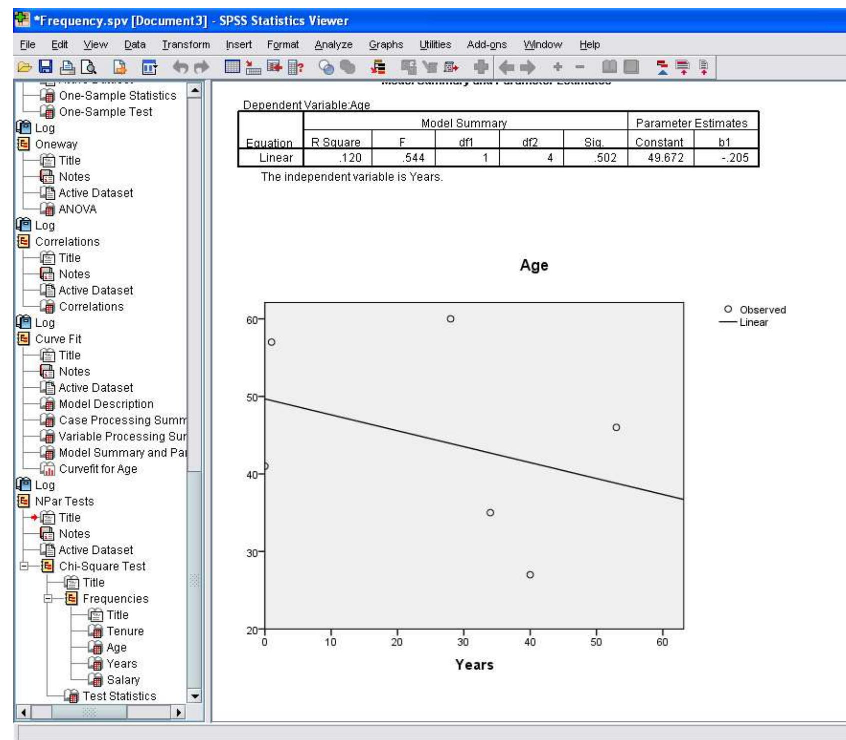
Curve Estimation

The Curve Estimation method generates curve estimation regression statistics and related plots for 11 different curve estimation regression models. A separate model is constructed for each dependent variable. You can also save predicted values, residuals and prediction intervals as new variables.



NOTES





Partial Least Squares Regression

The Partial Least Squares Regression procedure estimates Partial Least Squares (PLS, also known as 'Projection to Latent Structure') regression models. PLS is a predictive technique that is an alternative to Ordinary Least Squares (OLSs) regression, canonical correlation or structural equation modeling and it is particularly useful when predictor variables are highly correlated or when the number of predictors exceeds the number of cases.

PLS combines features of principal components analysis and multiple regression. It first extracts a set of latent factors that explain as much of the

covariance as possible between the independent and dependent variables. Then a regression step predicts values of the dependent variables using the decomposition of the independent variables.

NOTES

Availability: PLS is an extension command that requires the Python Extension Module to be installed on the system where you plan to run PLS. The PLS Extension Module is dependent upon Python software.

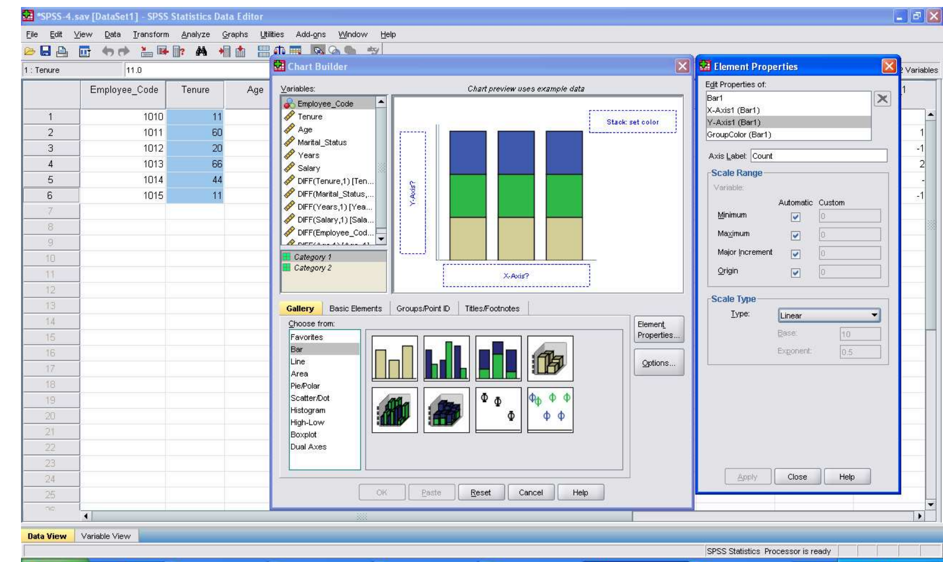
Tables: Proportion of variance explained (by latent factor), latent factor weights, latent factor loadings, independent Variable Importance in Projection (VIP) and regression parameter estimates (by dependent variable) are all produced by default.

Charts: Variable Importance in Projection (VIP), factor scores, factor weights for the first three latent factors and distance to the model are all produced from the Options tab.

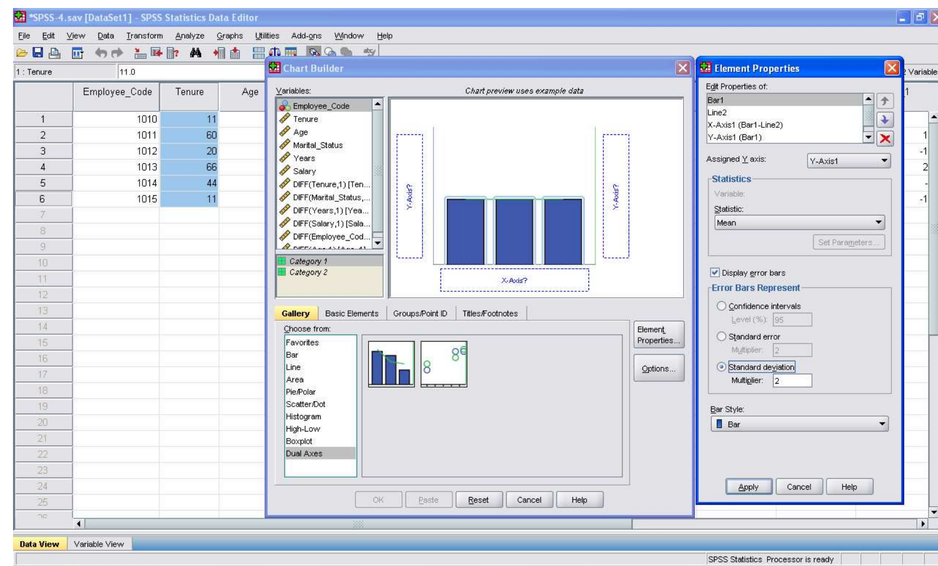
12.11 BUILDING CHARTS

The Chart Builder allows you to build charts from predefined gallery charts or from the individual parts (for example, axes and bars). You build a chart by dragging and dropping the gallery charts or basic elements onto the canvas, which is the large area to the right of the Variables list in the Chart Builder dialog box. As you are building the chart, the canvas displays a preview of the chart. Although the preview uses defined variable labels and measurement levels, it does not display your actual data. Instead, it uses randomly generated data to provide a rough sketch of how the chart will look.

Using the gallery is the preferred method for new users. You can also build a chart from basic elements. This is a more complex method because the gallery charts have predefined options that you need to define explicitly when building from the basic elements.



NOTES



12.11.1 Chart Builder Layout and Terms

The following are the properties and terms of Chart Builder:

Canvas

The canvas is the area of the Chart Builder dialog box where you build the chart.

Axis Set

An axis set defines one or more axes in a particular coordinate space, such as 2-D rectangular or 1-D polar. Adding a gallery item to the canvas automatically creates an axis set. You can also choose one from the Basic Elements tab. Each axis includes an axis drop zone for dragging and dropping variables. Blue text indicates that the zone still requires a variable. Every chart requires that you add a variable to the X-axis drop zone.

Graphic Elements

The graphic elements are the items in the chart that represent data. These are the bars, points, lines and so on.

Variables List

The variables list displays the available variables. If a variable selected in this list is categorical, the Categories list shows the defined categories for the variable. If the Categories list is not visible then drag the bar that separates the categories and variables lists. Similarly, you can use the Categories list to view the variables that make up a multiple response set. You can also temporarily change a variable's measurement level by right clicking its name in the Variables list and choosing one of the measurement levels.

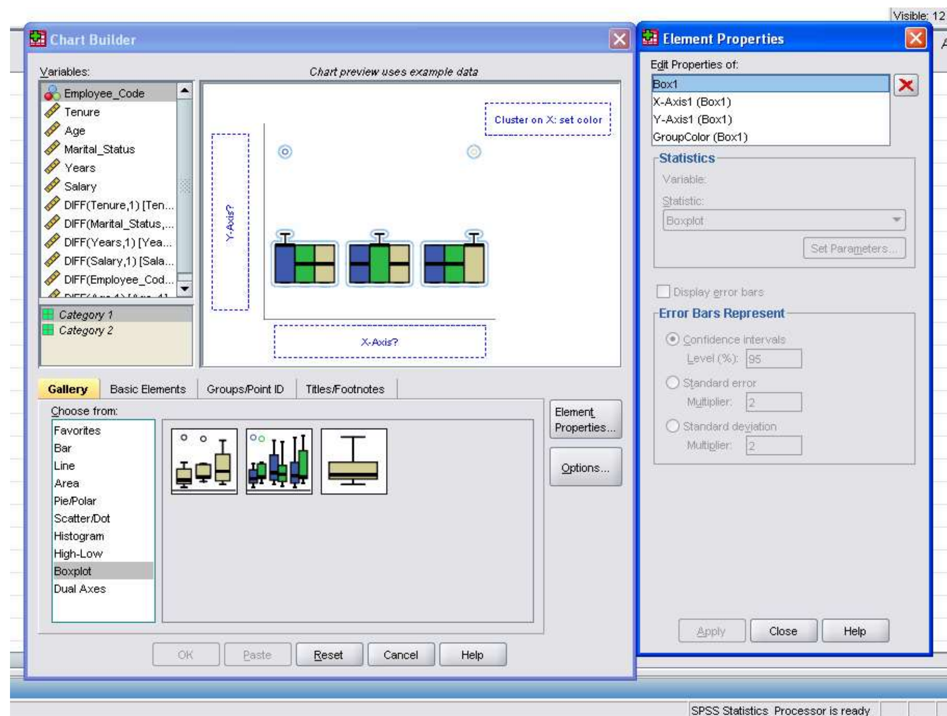
NOTES

Drop Zones

Drop zones are the areas on the canvas to which you drag and drop a variable from the Variables list. As noted previously, the basic drop zone is the axis drop zone. Certain gallery charts, such as clustered or stacked bar charts include grouping drop zones. You can also add these drop zones, paneling drop zones and point labeling drop zones from the Groups/Point ID tab.

Variable Types

The Chart Builder distinguishes the different measurement levels and handles the variables differently depending on the measurement level. In addition, the Chart Builder can graph multiple response sets which it treats as a categorical variable. An icon next to each variable in the Variables list identifies the variable type.



Measurement Level

A variable's measurement level is important when you create a chart. Following is a description of the measurement levels. You can temporarily change the measurement level in the Chart Builder by right clicking the variable in the Variables list and choosing an option. You can also permanently change a variable's measurement level in the Variable View of the Data Editor.

Categorical: Data with a limited number of distinct values or categories (for example, gender or religion). Categorical variables can be string (alphanumeric)

NOTES

or numeric variables that use numeric codes to represent categories (for example, 0 = Male and 1 = Female). Categorical variables can be either nominal or ordinal.

- **Nominal:** A variable can be treated as nominal when its values represent categories with no intrinsic ranking, for example, the department of the company in which an employee works. Examples of nominal variables include region, zip code and religious affiliation.
- **Ordinal:** A variable can be treated as ordinal when its values represent categories with some intrinsic ranking, for example, levels of service satisfaction from highly dissatisfied to highly satisfied. Examples of ordinal variables include attitude scores representing degree of satisfaction or confidence and preference rating scores.

Scale: Data measured on an interval or ratio scale, where the data values indicate both the order of values and the distance between values.

Categorical variables define categories in the chart, typically to draw separate graphic elements or to group graphic elements. Scale variables are often summarized within categories of categorical variables. For example, a default chart of income for gender categories would display the mean income for males and the mean income for females. The raw values for scale variables can also be plotted as it is done in a scatterplot. For example, a scatterplot may show the current salary and beginning salary for each case. A categorical variable could be used to group the cases by gender.

Defined Categories and Labels

A variable's defined categories are displayed in the Categories list and on the canvas when you use the categorical variable in a chart. If the variable has no defined categories, the Categories list and the canvas pane will display two placeholder categories: Category 1 and Category 2. The defined categories displayed in the Chart Builder are based on value labels, descriptive labels assigned to different data values, for example, numeric values of 0 and 1, with value labels of male and female respectively. You can define value labels in Variable View of the Data Editor or with Define Variable Properties on the Data menu in the Data Editor window.

Multiple Response Sets

Custom Tables and the Chart Builder support a special kind of 'variable' called a multiple response set. Multiple response sets aren't really 'variables' in the normal sense. You cannot see them in the Data Editor and other procedures do not recognize them. Multiple response sets use multiple variables to record responses to questions where the respondent can give more than one answer. Multiple response sets are

treated like categorical variables, and most of the things you can do with categorical variables, you can also do with multiple response sets.

Multiple response sets are constructed from multiple variables in the data file. A multiple response set is a special construct within a data file. You can define and save multiple response sets in SPSS Statistics data files, but you cannot import or export multiple response sets from/to other file formats. You can copy multiple response sets from other SPSS Statistics data files using Copy Data Properties, which is accessed from the Data menu in the Data Editor window.

NOTES

Building a Chart from the Gallery

The easiest method for building charts is to use the gallery. Following are general steps for building a chart from the gallery.

- Click the Gallery tab if it is not already displayed.
- In the Choose From list, select a category of charts. Each category offers several types.
- Drag the picture of the desired chart onto the canvas. You can also double-click the picture. If the canvas already displays a chart, the gallery chart replaces the axis set and graphic elements on the chart.
- Drag variables from the Variables list and drop them into the axis drop zones and, if available, the grouping drop zone. If an axis drop zone already displays a statistic and you want to use that statistic, you do not have to drag a variable into the drop zone. You need to add a variable to a zone only when the text in the zone is blue. If the text is black, the zone already contains a variable or statistic.

Note: The measurement level of your variables is important. The Chart Builder sets defaults based on the measurement level while you are building the chart. Furthermore, the resulting chart may also look different for different measurement levels. You can temporarily change a variable's measurement level by right-clicking the variable and choosing an option.

- If you need to change statistics or modify attributes of the axes or legends (such as the scale range), click Element Properties.
- In the Edit Properties Of list, select the item you want to change. (For information about the specific properties, click Help.)
- After making any changes, click Apply.
- If you need to add more variables to the chart (for example, for clustering or paneling), click the Groups/Point ID tab in the Chart Builder dialog box

NOTES

and select one or more options. Then drag categorical variables to the new drop zones that appear on the canvas.

- If you want to transpose the chart (for example, to make the bars horizontal), click the Basic Elements tab and then click Transpose.
- Click OK to create the chart. The chart is displayed in the Viewer.

If you often change many default settings for a specific chart, you may want to save it as a favorite.

Summarizing Separate Variables

You can use the Chart Builder to create charts that summarize and compare multiple variables. For example, if sales data for each year are recorded in a separate variable, you can create a single chart that summarizes sales data for each year in a separate bar.

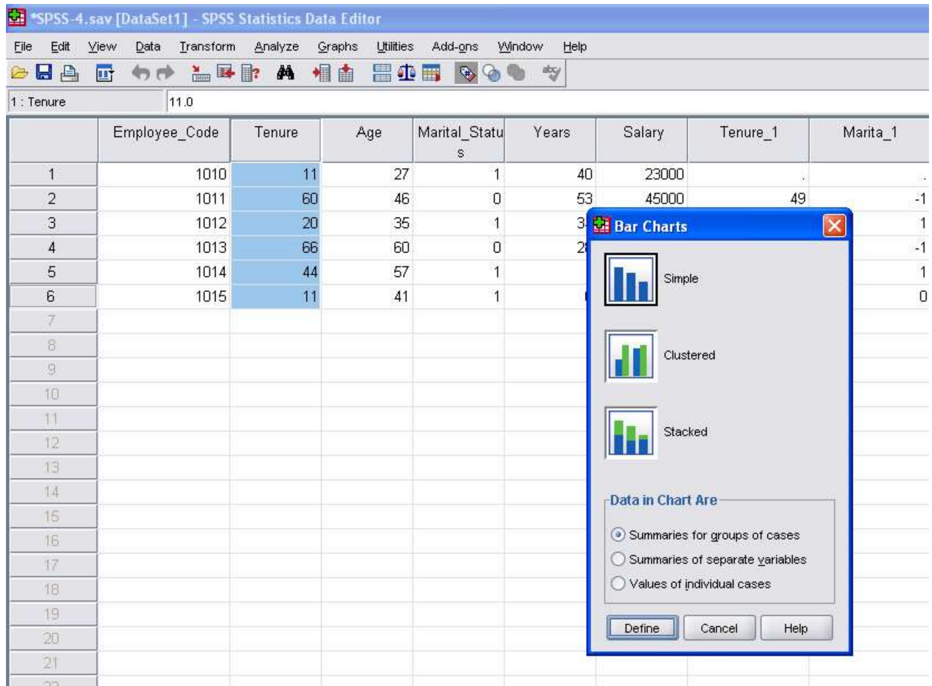
12.11.2 Chart Types

The gallery contains a collection of the most commonly used charts. These include:

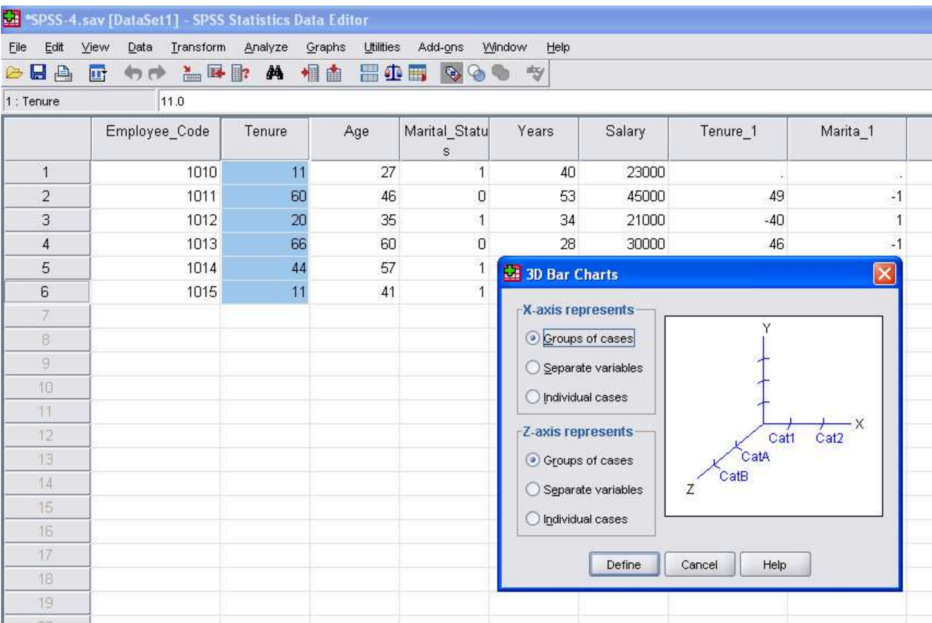
- Bar charts: Simple, stacked, clustered, 3-D, and error bar.
- Line charts: Simple and grouped (multi-line).
- Area charts: Simple and stacked.
- Pie charts: Simple.
- Scatterplots and dot plots: 1-D, simple, grouped, overlay, and 3-D scatterplots; summary point plots, 1-D dot plots and drop-line charts.
- Histograms: Simple, stacked, frequency polygons and population pyramids.
- High-Low charts: High-low-close, range bar, clustered range bar and difference area.
- Boxplots: Simple and clustered.
- Dual Y-axis charts: Simple.

The following are some examples of types of graphs as they look alike.

Bar Chart



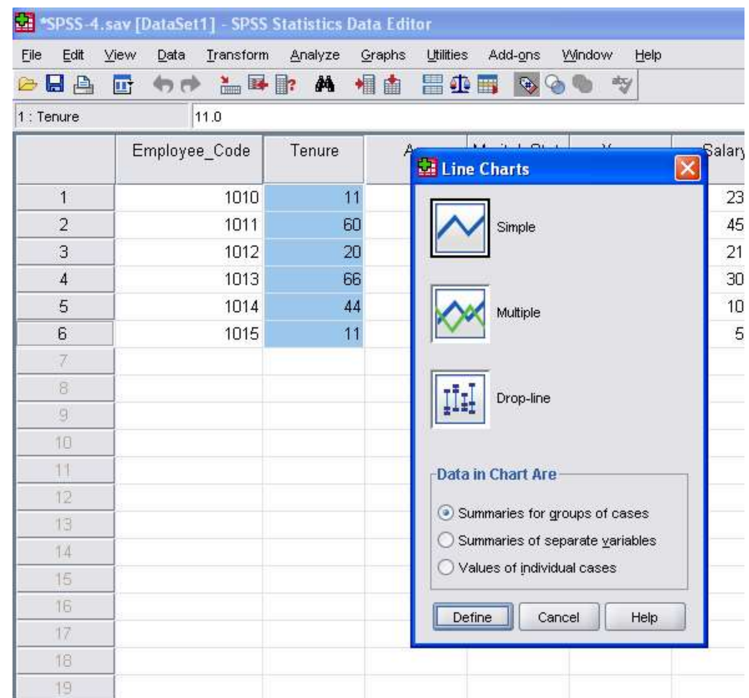
3-D Bar Chart



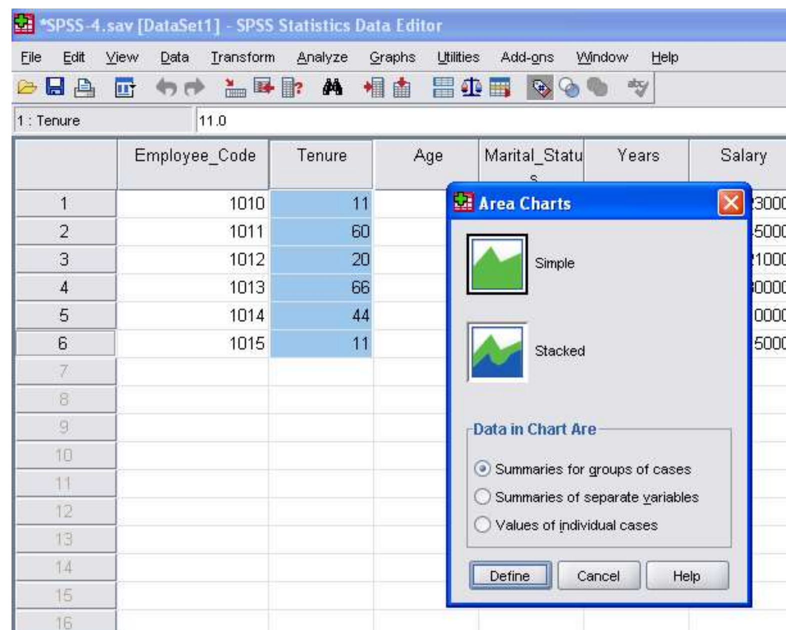
NOTES

NOTES

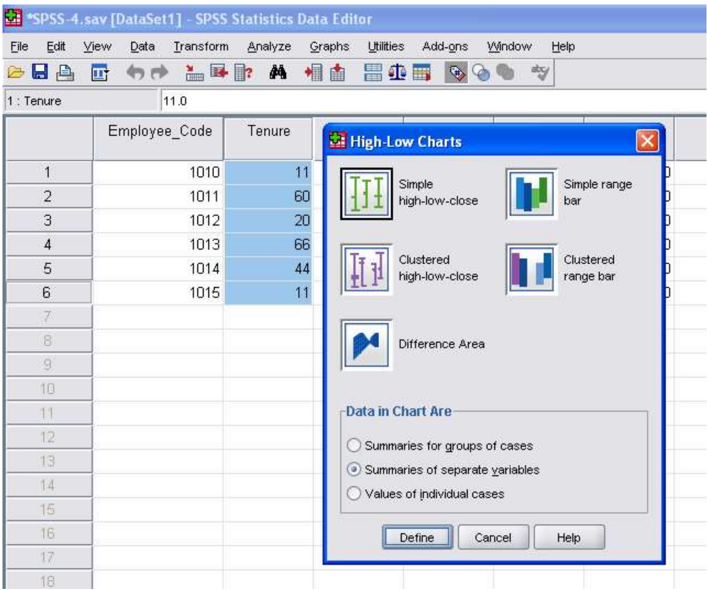
Line Graph



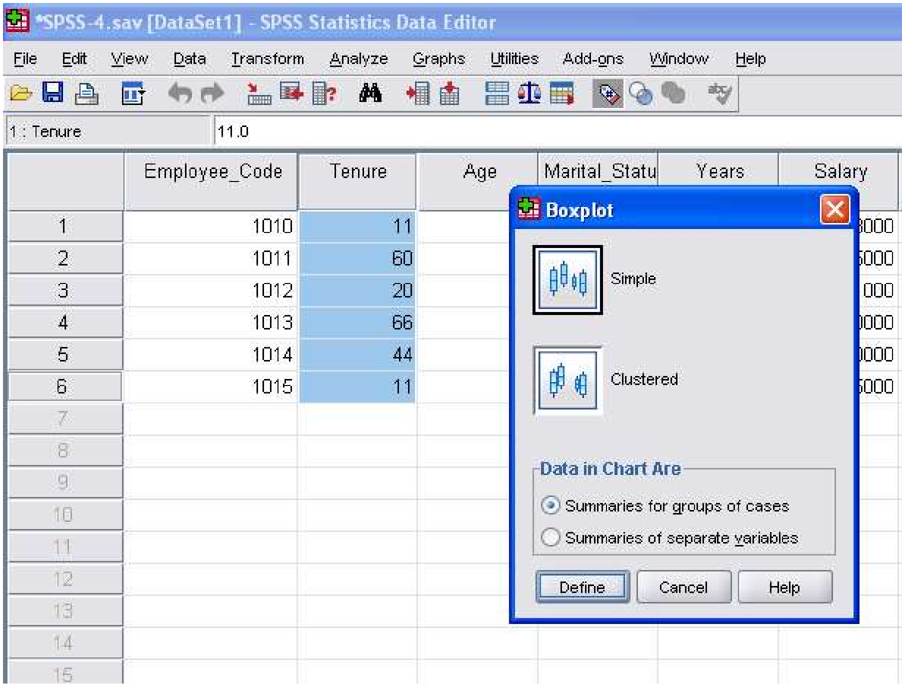
Area Chart



High-Low Charts



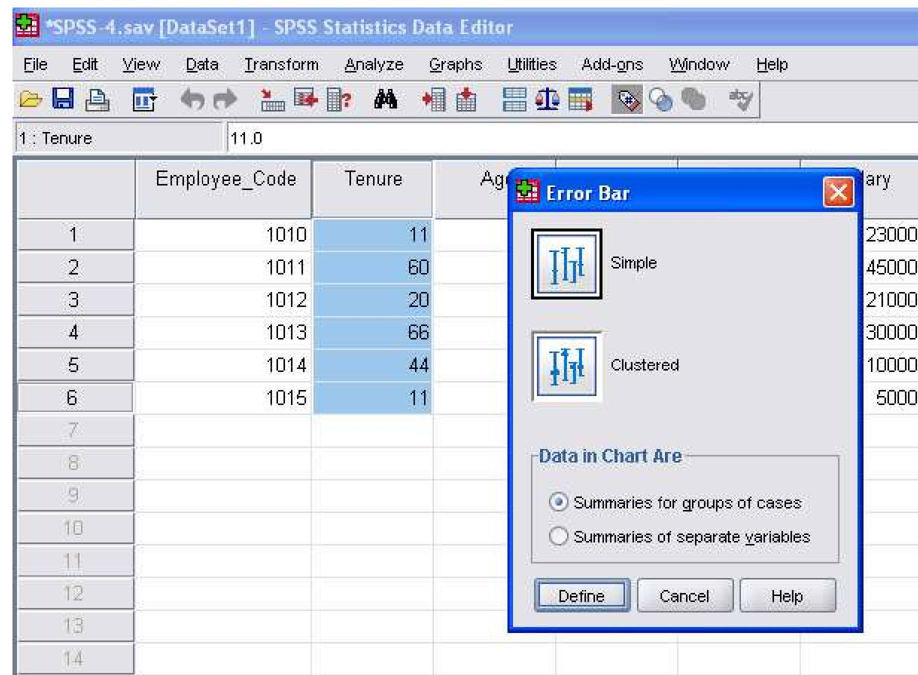
Boxplot Chart



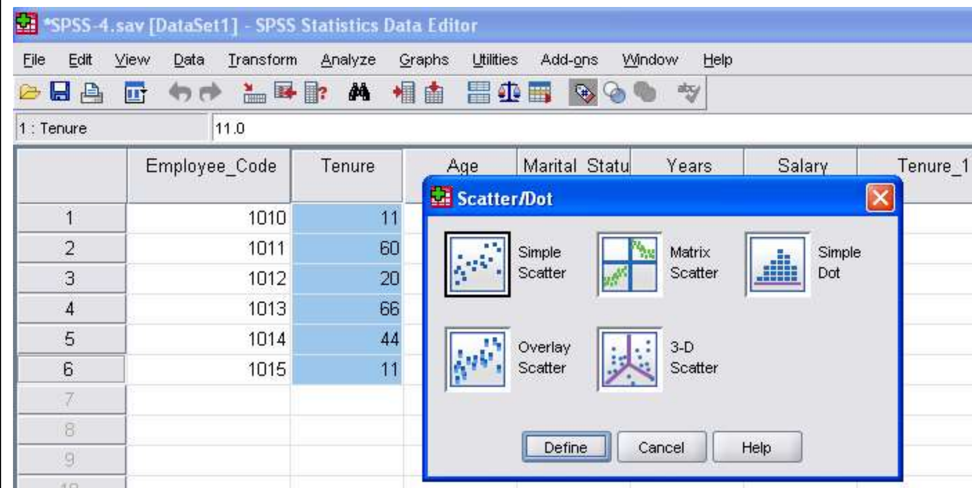
NOTES

NOTES

Error Bar Chart

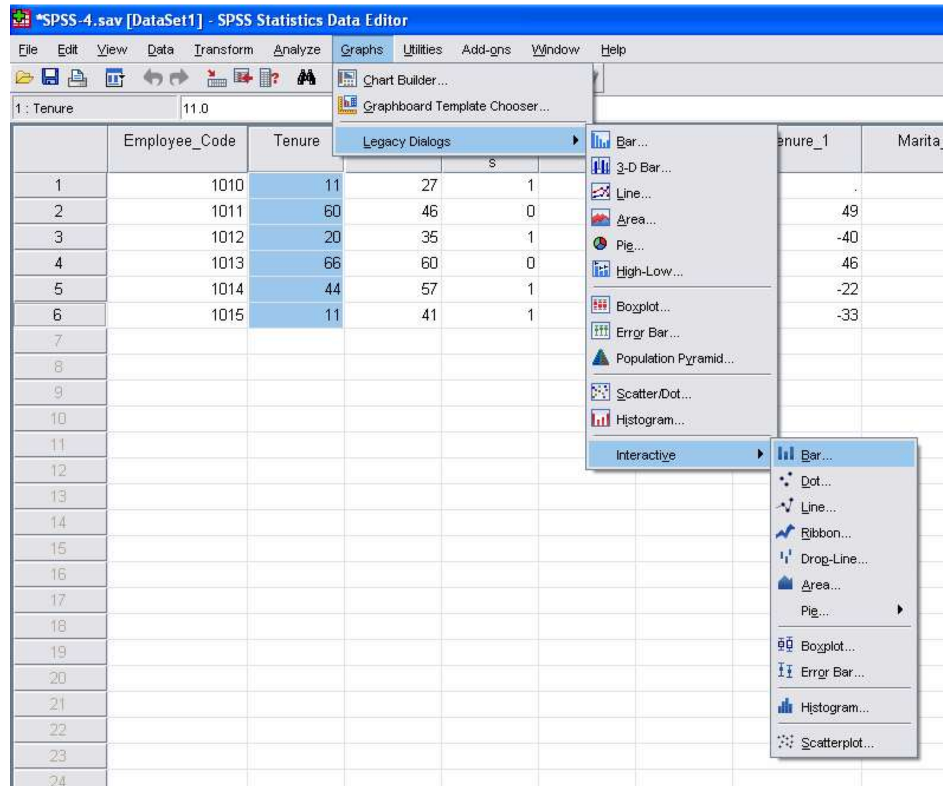


Scatter/Dot Chart



You can also draw interactive charts as shown in the following screen:

NOTES



Building a Chart from the Basic Elements

You can use the basic elements to build a chart from scratch. Basic elements include axes and graphic elements. These elements are 'basic' because you cannot create a chart without them. If you are using the Chart Builder for the first time, it is recommended that you use the gallery charts instead. Gallery charts automatically set properties and add features to simplify the creation process.

Click the Basic Elements tab if it is not already displayed. If the canvas is empty, drag an axis set onto the canvas.

Add a graphic element type by dragging a graphic element onto the canvas. Not all graphic elements can be used with a particular axis set. Only the relevant graphic elements are enabled.

Drag variables from the Variables list and drop them into the axis drop zones and, if available, the grouping drop zone. If an axis drop zone already displays a statistic and you want to use that statistic, you do not have to drag a variable into the drop zone. You need to add a variable to a zone only when the text in the zone is blue. If the text is black, the zone already contains a variable or statistic.

Note: The measurement level of your variables is important.

NOTES

The Chart Builder sets defaults based on the measurement level while you are building the chart. Furthermore, the resulting chart may also look different for different measurement levels. You can temporarily change a variable's measurement level by right-clicking the variable and choosing an option.

If you need to change statistics or modify attributes of the axes or legends (such as the scale range), click Element Properties.

In the Edit Properties of list, select the item you want to change. For information about the specific properties, click Help.

After making any changes, click Apply.

If you need to add more variables to the chart (for example, for clustering or paneling), click the Groups/Point ID tab in the Chart Builder dialog box and select one or more options. Then drag categorical variables to the new drop zones that appear on the canvas.

If you want to transpose the chart (for example, to make the bars horizontal), click Transpose.

Click OK to create the chart. The chart is displayed in the Viewer.

If you often create a customized chart, you may want to save it as a favorite chart so you do not have to recreate it from scratch each time.

Using Favorites

You can save all the settings associated with a chart so that you do not have change the defaults each time you create a particular chart. Everything that you can modify with the Chart Builder is saved with the favorite, including statistics and axis properties such as the range. Variables are not saved in the favorite. The favorite is available until you manually delete it.

Editing Elements

A chart element is a graphic element (for example, a bar), an axis, a grouping zone, or one of the optional elements (such as titles and footnotes). In other words, elements are all of the items that you add to the chart. You can edit these elements by modifying their properties. For example, you can change the statistic that controls the bar height, reorder of the categories on a categorical axis, or transform the axis scale.

To Modify the Element Properties

Click Element Properties to display the Element Properties dialog box. You use the same dialog box to modify all elements in the chart.

In the Edit Properties Of list, select the element that you want to modify.

Click Apply to save your changes. If you select a different element before clicking Apply, your changes will be discarded. You must save your changes before editing another element.

Check the related topics for information about specific properties.

NOTES

Editing Charts

The Chart Editor provides a powerful, easy-to-use environment where you can customize your charts and explore your data. The Chart Editor features:

- Simple, intuitive user interface. You can quickly select and edit parts of the chart using menus, context menus, and toolbars. You can also enter text directly on a chart.
- Wide range of formatting and statistical options. You can choose from a full range of styles and statistical options.
- Powerful exploratory tools. You can explore your data in various ways, such as by labeling, reordering, and rotating it. You can change chart types and the roles of variables in the chart. You can also add distribution curves and fit, interpolation, and reference lines.
- Flexible templates for consistent look and behavior. You can create customized templates and use them to easily create charts with the look and options that you want. For example, if you always want a specific orientation for axis labels, you can specify the orientation in a template and apply the template to other charts.

Chart Options

Chart Template: New charts can use either the settings selected here or the settings from a chart template file. Click Browse to select a chart template file. To create a chart template file, create a chart with the attributes that you want and save it as a template (choose Save Chart Template from the File menu).

Chart Aspect Ratio: The width-to-height ratio of the outer frame of new charts. You can specify a width-to-height ratio from 0.1 to 10.0. Values less than 1 make charts that are taller than they are wide. Values greater than 1 make charts that are wider than they are tall. A value of 1 produces a square chart. Once a chart is created, its aspect ratio cannot be changed.

Current Settings: Available settings include:

- **Font:** Font used for all text in new charts.
- **Style Cycle Preference:** The initial assignment of colors and patterns for new charts. Cycle through colors only uses only colors to differentiate chart elements and does not use patterns. Cycle through patterns only uses only line styles, marker symbols or fill patterns to differentiate chart elements and does not use color.
- **Frame:** Controls the display of inner and outer frames on new charts.
- **Grid Lines:** Controls the display of scale and category axis grid lines on new charts.

Style Cycles: Customizes the colors, line styles, marker symbols, and fill patterns for new charts. You can change the order of the colors and patterns that are used when a new chart is created.

NOTES

12.12 VIEWER OPTIONS FOR OUTPUT

Viewer output display options affect only new output produced after you change the settings. Output already displayed in the Viewer is not affected by changes in these settings.

Initial Output State: Controls which items are automatically displayed or hidden each time you run a procedure and how items are initially aligned. You can control the display of the following items: log, warnings, notes, titles, pivot tables, charts, tree diagrams, and text output. You can also turn the display of commands in the log on or off. You can copy command syntax from the log and save it in a syntax file.

Note: All output items are displayed left-aligned in the Viewer. Only the alignment of printed output is affected by the justification settings. Centered and right-aligned items are identified by a small symbol above and to the left of the item.

Title: Controls the font style, size, and color for new output titles.

Page Title: Controls the font style, size, and color for new page titles and page titles generated by TITLE and SUBTITLE command syntax or created by New Page Title on the Insert menu.

Text Output Font Used for Text Output: Text output is designed for use with a monospaced (fixed-pitch) font. If you select a proportional font, tabular output will not align properly.

Viewer

Results are displayed in the Viewer. You can use the Viewer to:

- Browse results.
- Show or hide selected tables and charts.
- Change the display order of results by moving selected items.
- Move items between the Viewer and other applications.

The Viewer is divided into two panes:

- The left pane contains an outline view of the contents.
- The right pane contains statistical tables, charts and text output.

You can click an item in the outline to go directly to the corresponding table or chart. You can click and drag the right border of the outline pane to change the width of the outline pane.

NOTES

Showing and Hiding Results

In the Viewer, you can selectively show and hide individual tables or results from an entire procedure. This process is useful when you want to shorten the amount of visible output in the contents pane.

To Hide Tables and Charts

Double click the item's book icon in the outline pane of the Viewer or Click the item to select it. From the menu select **View → Hide.**

To Hide Procedure Results

Click the box to the left of the procedure name in the outline pane. This hides all results from the procedure and collapses the outline view.

Moving, Deleting, and Copying Output

You can rearrange the results by copying, moving or deleting an item or a group of items.

To Move Output in the Viewer

Select the items in the outline or contents pane. Drag and drop the selected items into a different location.

To Delete Output in the Viewer

Select the items in the outline or contents pane. Press the Delete key or from the menu select **Edit → Delete.**

Changing Initial Alignment

By default, all results are initially left aligned. To change the initial alignment of new output items select the alignment option you want.

Changing Alignment of Output Items

In the outline or contents pane, select the items that you want to align.

From the menu select **Format → Align Left or Format → Center or Format → Align Right.**

Viewer Outline

The outline pane provides a table of contents of the Viewer document. You can use the outline pane to navigate through your results and control the display. Most actions in the outline pane have a corresponding effect on the contents pane.

- Selecting an item in the outline pane displays the corresponding item in the contents pane.

- Moving an item in the outline pane moves the corresponding item in the contents pane.
- Collapsing the outline view hides the results from all items in the collapsed levels.

NOTES

Controlling the Outline Display: To control the outline display, you can,

- Expand and collapse the outline view.
- Change the outline level for selected items.
- Change the size of items in the outline display.
- Change the font that is used in the outline display.

To Collapse and Expand the Outline View

Click the box to the left of the outline item that you want to collapse or expand.

Adding Items to the Viewer

In the Viewer, you can add items such as titles, new text, charts or material from other applications.

To Add a Title or Text

Text items that are not connected to a table or chart can be added to the Viewer. Click the table, chart or other object that will precede the title or text.

To Add a Text File

In the outline pane or contents pane of the Viewer, click the table, chart or other object that will precede the text. To edit the text, double click it.

Pasting Objects into the Viewer

Objects from other applications can be pasted into the Viewer. You can use either Paste After or Paste Special. Either type of pasting puts the new object after the currently selected object in the Viewer. Use Paste Special when you want to choose the format of the pasted object.

Finding and Replacing Information in the Viewer

To find or replace information in the Viewer, from the main menu select **Edit** → **Find**.

You can use Find and Replace to:

- Search the entire document or just the selected items.
- Search down or up from the current location.
- Search both panes or restrict the search to the contents or outline pane.

- Search for hidden items. These include any items hidden in the contents pane (for example, Notes tables, which are hidden by default) and hidden rows and columns in pivot tables.
- Restrict the search criteria to case sensitive matches.
- Restrict the search criteria in pivot tables to matches of the entire cell contents.

NOTES

Hidden Items and Pivot Table Layers

- Layers beneath the currently visible layer of a multidimensional pivot table are not considered hidden and will be included in the search area even when hidden items are not included in the search.
- Hidden items include hidden items in the contents pane (items with closed book icons in the outline pane or included within collapsed blocks of the outline pane) and rows and columns in pivot tables either hidden by default (for example, empty rows and columns are hidden by default) or manually hidden by editing the table and selectively hiding specific rows or columns. Hidden items are only included in the search if you explicitly select Include hidden items.
- In both cases, the hidden or nonvisible element that contains the search text or value is displayed when it is found, but the item is returned to its original state afterward.

Copying Output into Other Applications

Output objects can be copied and pasted into other applications, such as a word-processing program or a spreadsheet. You can paste output in a variety of formats. Depending on the target application, some or all of the following formats may be available:

Picture (Metafile): Pivot tables and charts can be pasted as metafile pictures. The picture format can be resized in the other application, and sometimes a limited amount of editing can be done with the facilities of the other application. Pivot tables that are pasted as pictures retain all borders and font characteristics. For pivot tables, select Files from the Paste Special menu. This format is available only on Windows operating systems.

RTF (Rich Text Format): Pivot tables can be pasted into other applications in RTF format. In most applications, this means that the pivot table is pasted as a table that can then be edited in the other application.

Note: Microsoft Word may not display extremely wide tables properly.

Bitmap: Charts and other graphics can be pasted into other applications as bitmaps.

BIFF: The contents of a table can be pasted into a spreadsheet and retain numeric precision. This format is available only on Windows operating systems.

Text: The contents of a table can be copied and pasted as text. This process can be useful for applications such as e-mail, where the application can accept or transmit only text.

NOTES

If the target application supports multiple available formats, it may have a Paste Special menu item that allows you to select the format, or it may automatically display a list of available formats.

Export Output

Export Output saves Viewer output in HTML, text, Word/RTF, Excel, PowerPoint (requires PowerPoint 97 or later), and PDF formats. Charts can also be exported in a number of different graphics formats.

Note: Export to PowerPoint is available only on Windows operating systems and is not available in the Student Version.

Objects to Export: You can export all objects in the Viewer, all visible objects, or only selected objects.

Document Type: The available options are:

- **Word/RTF (*.doc):** Pivot tables are exported as Word tables with all formatting attributes intact—for example, cell borders, font styles, and background colors. Text output is exported as formatted RTF. Charts, tree diagrams, and model views are included in PNG format.

  Note: Microsoft Word may not display extremely wide tables properly.

- **Excel (*.xls):** Pivot table rows, columns, and cells are exported as Excel rows, columns, and cells, with all formatting attributes intact—for example, cell borders, font styles, and background colors. Text output is exported with all font attributes intact. Each line in the text output is a row in the Excel file, with the entire contents of the line contained in a single cell. Charts, tree diagrams, and model views are included in PNG format.
- **HTML (*.htm):** Pivot tables are exported as HTML tables. Text output is exported as preformatted HTML. Charts, tree diagrams, and model views are embedded by reference, and you should export charts in a suitable format for inclusion in HTML documents (for example, PNG and JPEG).
- **Portable Document Format (*.pdf):** All output is exported as it appears in Print Preview, with all formatting attributes intact.
- **PowerPoint file (*.ppt):** Pivot tables are exported as Word tables and are embedded on separate slides in the PowerPoint file, with one slide for each pivot table. All formatting attributes of the pivot table are retained—for example, cell borders, font styles, and background colors. Charts, tree diagrams, and model views are exported in TIFF format. Text output is not included.

  Export to PowerPoint is available only on Windows operating systems.

NOTES

- **Text (*.txt):** Text output formats include plain text, UTF-8, and UTF-16. Pivot tables can be exported in tab-separated or space-separated format. All text output is exported in space-separated format. For charts, tree diagrams, and model views, a line is inserted in the text file for each graphic, indicating the image filename.
- **None (Graphics Only):** Available export formats include: EPS, JPEG, TIFF, PNG, and BMP. On Windows operating systems, EMF (enhanced metafile) format is also available.

Output Management System. You can also automatically export all output or user-specified types of output as Word, Excel, PDF, HTML, text or SPSS Statistics-format data files.

Viewer Printing

There are two options for printing the contents of the Viewer window:

All Visible Output: Prints only items that are currently displayed in the contents pane. Hidden items (items with a closed book icon in the outline pane or hidden in collapsed outline layers) are not printed.

Selection: Prints only items that are currently selected in the outline and/or contents panes.

Saving Output

The contents of the Viewer can be saved to a Viewer document. The saved document includes both panes of the Viewer window (the outline and the contents).

Check Your Progress

1. How the variables are transformed in data?
2. In SPSS can the user modify the window displays and also the output format?
3. How the data type for a variable can be changed in SPSS?
4. Write the steps for printing the data file in SPSS.
5. Explain the Scatter diagram method.
6. Why is 'Curve Estimation' method used?
7. Explain the features of 'Chart Builder' method used in SPSS.

12.13 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. You can Transform or combine variables to create new variables. The Compute and Recode commands that are available in the Transform pull

NOTES

down menu help you to edit the data files. Thus with the Data Editor, you can modify data values in Data View using the following options:

- Change data values
- Cut, copy and paste data values
- Add and delete cases
- Add and delete variables
- Change the order of variables

2. There are numerous options available in SPSS that permit the user to modify the window displays and also the output format. These options can be accessed by selecting from the main menu Edit ® Options. Check that variable lists show the name of the variable and not its label. You can also change the font options and other output display characteristics.
3. The data type for a variable can be changed at any time by using the Variable Type dialog box in Variable View. Using the Data Editor you can convert the existing values to the new type. The system missing value is automatically allocated if no conversion is feasible. All the conversion rules are identical for entering or pasting data values to a variable with a different format type. The modification in data format can result in the failure of missing-value specifications or value labels. The Data Editor shows an alert box and inquires whether the changes should be accepted or cancelled before you proceed further.
4. A data file can be printed in the form as it appears on the screen. The following are the printing specifications of data file in Data Editor:
   - You can print the information that is currently displayed. You can print the data in Data View where as in Variable View you can only print the data definition information.
   - You can print the Grid lines if they are currently displayed in the selected view.
   - Value labels are printed in Data View if they are currently displayed. Otherwise, the actual data values are printed.
   - Use the View menu in the Data Editor window to display or hide grid lines and toggle between the display of data values and value labels.
5. Scatter diagram method makes use of the Scatter diagram also known as Dot diagram. Scatter diagram is a diagram representing two series with the known variable, i.e., independent variable plotted on the X-axis and the variable to be estimated, i.e., dependent variable to be plotted on the Y-axis on a graph paper.
6. The 'Curve Estimation' method generates curve estimation regression statistics and related plots for 11 different curve estimation regression models.

NOTES

A separate model is constructed for each dependent variable. You can also save predicted values, residuals and prediction intervals as new variables.

7. The 'Chart Builder' allows to build charts from predefined gallery charts or from the individual parts (for example, axes and bars). You build a chart by dragging and dropping the gallery charts or basic elements onto the canvas, which is the large area to the right of the Variables list in the Chart Builder dialog box.

   As you are building the chart, the canvas displays a preview of the chart. Although the preview uses defined variable labels and measurement levels, it does not display your actual data. Instead, it uses randomly generated data to provide a rough sketch of how the chart will look.

   A chart element is a graphic element (for example, a bar), an axis, a grouping zone, or one of the optional elements (such as titles and footnotes).

   Elements are all of the items that you add to the chart. You can edit these elements by modifying their properties. For example, you can change the statistic that controls the bar height, reorder of the categories on a categorical axis, or transform the axis scale.

   The Chart Editor provides a powerful, easy-to-use environment where the charts are customized.

12.14 SUMMARY

- In Data View, you can enter data directly in the Data Editor in any sequence. Data can be entered either by case or by variable specifically for selected areas or for individual cells.
- When a value is entered in a blank column then a new variable is automatically created. The Data Editor assigns the variable name.
- You can select the cases by using 'Selected' option. Selection is either temporary or permanent. Temporary selection is done using 'Filtered' option which means that the complete dataset remains in memory even though a specific subset is being analysed whereas permanent selection is done using 'Deleted' option which means that the complete dataset is deleted and not available for future use.
- You can sort the dataset by using 'Sorted' option. Sorting is done according to the values of one or more variables.
- You can merge files by using the 'Merged' option. Merging is done in two different ways, either by adding new cases (rows) to an existing file using the 'Add Cases' option or by adding variables (columns) to the existing file using the 'Add Variables' option.

NOTES

- You can Transform or combine variables to create new variables. The Compute and Recode commands that are available in the Transform pull down menu help you to edit the data files.
- You can Transform or combine variables to create new variables. The Compute and Recode commands that are available in the Transform pull down menu help you to edit the data files. Thus with the Data Editor, you can modify data values in Data View using the following options:
  (i) Change data values
  (ii) Cut, copy and paste data values
  (iii) Add and delete cases
  (iv) Add and delete variables
  (v) Change the order of variables
- There are numerous options available in SPSS that permit the user to modify the window displays and also the output format. These options can be accessed by selecting from the main menu Edit ® Options. Check that variable lists show the name of the variable and not its label. You can also change the font options and other output display characteristics.
- Generally, SPSS produces discrete output objects rather than ordinary ASCII text. These output objects are immensely formatted tables and are termed as Pivot Tables or Charts which can be easily edited within SPSS.
- The data type for a variable can be changed at any time by using the Variable Type dialog box in Variable View.
- Using the Data Editor you can convert the existing values to the new type. The system missing value is automatically allocated if no conversion is feasible. All the conversion rules are identical for entering or pasting data values to a variable with a different format type.
- The modification in data format can result in the failure of missing-value specifications or value labels. The Data Editor shows an alert box and inquires whether the changes should be accepted or canceled before you proceed further.
- A data file can be printed in the form as it appears on the screen.
- You can print the information that is currently displayed. You can print the data in Data View where as in Variable View you can only print the data definition information.
- You can print the Grid lines if they are currently displayed in the selected view.
- Value labels are printed in Data View if they are currently displayed. Otherwise, the actual data values are printed.

NOTES

- Use the View menu in the Data Editor window to display or hide grid lines and toggle between the display of data values and value labels.
- You can copy both data and variable definition attributes from one dataset to another dataset in basically the same way that you copy and paste information within a single data file.
- You can open a data file or enter data in the Data Editor and then can start creating reports, charts and analyses without any additional preliminary work.
- Cases are considered duplicates if their values match for all selected variables. If you want to identify only cases that are a 100% match in all respects, select all of the variables. Sort within matching groups by.
- Cases are automatically sorted by the variables that define matching cases. You can select additional sorting variables that will determine the sequential order of cases in each matching group.
- Scatter diagram method makes use of the Scatter diagram also known as Dot diagram. Scatter diagram is a diagram representing two series with the known variable, i.e., independent variable plotted on the X-axis and the variable to be estimated, i.e., dependent variable to be plotted on the Y-axis on a graph paper.
- The 'Curve Estimation' method generates curve estimation regression statistics and related plots for 11 different curve estimation regression models. A separate model is constructed for each dependent variable. You can also save predicted values, residuals and prediction intervals as new variables.
- The 'Chart Builder' allows to build charts from predefined gallery charts or from the individual parts (for example, axes and bars). You build a chart by dragging and dropping the gallery charts or basic elements onto the canvas, which is the large area to the right of the Variables list in the Chart Builder dialog box.
- As you are building the chart, the canvas displays a preview of the chart. Although the preview uses defined variable labels and measurement levels, it does not display your actual data. Instead, it uses randomly generated data to provide a rough sketch of how the chart will look.
- Using the gallery is the preferred method for new users. You can also build a chart from basic elements. This is a more complex method because the gallery charts have predefined options that you need to define explicitly when building from the basic elements.
- A chart element is a graphic element (for example, a bar), an axis, a grouping zone, or one of the optional elements (such as titles and footnotes).
- Elements are all of the items that you add to the chart. You can edit these elements by modifying their properties. For example, you can change the

NOTES

statistic that controls the bar height, reorder of the categories on a categorical axis, or transform the axis scale.

- The Chart Editor provides a powerful, easy-to-use environment where you can customize your charts and explore your data.

12.15 KEY WORDS

- **Data View:** In Data View, you can enter data directly in the Data Editor in any sequence. Data can be entered either by case or by variable specifically for selected areas or for individual cells.
- **Fonts:** This option controls the font characteristics of the data to be displayed.
- **Label:** It displays any value labels that have already been defined. You can add or change labels in this column.
- **Value:** It displays unique values for each selected variable. This list of unique values is based on the number of scanned cases.
- **Count:** It displays the number of times each value occurs in the scanned cases.
- **Scatter diagram:** Scatter diagram method makes use of the Scatter diagram also known as Dot diagram. Scatter diagram is a diagram representing two series with the known variable, i.e., independent variable plotted on the X-axis and the variable to be estimated, i.e., dependent variable to be plotted on the Y-axis on a graph paper.

12.16 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. What does output display mean?
2. How is numeric data entered in data view?
3. Why value labels are used for data entry?
4. Why is data edited?
5. Why are data values replaced or modified?
6. How are new cases inserted?
7. Why is data file printed?
8. Why are multiple data sources used?
9. What is the significance of data preparation?

10. What is value label grid?
11. What are multiple response sets?
12. What are dataset properties?
13. How charts are built in SPSS?
14. Name the various chart types available in SPSS.
15. Why is viewer options used for output?

NOTES

Long-Answer Questions

1. Briefly discuss the concept of output display, interpretation of output and display formats with the help of appropriate examples.
2. Describe the process of entering various types of data in data view with the help of examples.
3. Explain the process of data editing and the significance of display log in SPSS.
4. Explain the methodology of replacing or modifying data values with the help of examples.
5. Describe the process of data conversion in the data editor.
6. Explain the significance of inserting new cases and new variables in SPSS.
7. Describe the use of case selection in data editor? How is it done?
8. How is information displayed in data view and variable view printed? Explain with the help of examples.
9. Describe the significance and functioning of multiple data sources in SPSS.
10. Describe the role of variable properties in SPSS. How are variable properties defined?
11. How are value labels and other variable properties defined in SPSS? Explain.
12. Describe the significance and functioning of multiple response sets in SPSS.
13. Explain the process of identifying duplicate cases.
14. Briefly explain the process of data transformation.
15. Differentiate between scatter diagram and least squares method.
16. How is curve estimation done in SPSS? Explain with the help of an example.
17. Describe the importance of chart builder and also explain the properties and terms used in chart builder.
18. Explain the various types of charts used in SPSS with the help of examples.
19. What is the significance of viewer option in SPSS output format? Explain all the features and properties of viewer option with the help of examples.

12.17 FURTHER READINGS

NOTES

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

BLOCK - IV
COMPUTER APPLICATIONS TO ECONOMICS

NOTES

UNIT 13 APPLICATION TO ECONOMICS

Structure

13.0 INTRODUCTION

Economics is the social science that studies the production, distribution, and consumption of goods and services. Economics focuses on the behaviour and interactions of economic agents and how economies work. Microeconomics analyzes basic elements in the economy, including individual agents and markets, their interactions, and the outcomes of interactions. Individual agents may include, for example, households, firms, buyers, and sellers. E-commerce (Electronic commerce) is the activity of electronically buying or selling of products on online services or over the Internet. Electronic commerce draws on technologies, such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, online transaction processing, Electronic Data Interchange (EDI), inventory management systems, and automated data collection systems. Computational economics is a discipline at the interface of computer science, economics, and management science. This subject encompasses computational modelling of economic systems, whether macroeconomic, or rational-expectations, computational econometrics and statistics, computational finance, computational tools for the design of automated Internet markets, programming tool specifically designed for computational economics.

NOTES

In this unit, you will study about the computer applications to Economics, on-line banking, ATM's, electronic stock exchange, electronic trading, data sharing and dissemination, electronic transaction, document delivery, authentication and validation of transaction processing.

13.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the concept of computer application to Economics
- Explain about on-line banking and ATM's
- Define the electronic stock exchange and electronic trading
- Elaborate on data sharing and dissemination
- Discuss the electronic transaction, document delivery and authentication and validation transaction processing

13.2 ONLINE BANKING

Online Banking, also known as Internet Banking or Web Banking, is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website. The online banking system will typically connect to or be part of the core banking system operated by a bank and is in contrast to branch banking which was the traditional way customers accessed banking services.

Some banks operate as a 'Direct Bank' or 'Virtual Bank', where they rely completely on Internet banking. Internet banking software provides personal and corporate banking services offering features, such as viewing account balances, obtaining statements, checking recent transactions, transferring money between accounts, and making payments.

To access a financial institution's online banking facility, a customer with Internet access will need to register with the institution for the service, and set up a password and other credentials for customer verification. The credentials for online banking is normally not the same as for telephone or mobile banking. Financial institutions now routinely allocate customers numbers, whether or not customers have indicated an intention to access their online banking facility. Customer numbers are normally not the same as account numbers, because a number of customer accounts can be linked to the one customer number. Technically, the customer number can be linked to any account with the financial institution that the customer controls, though the financial institution may limit the range of accounts that may be accessed to, say, cheque, savings, loan, credit card and similar accounts.

The customer visits the financial institution's secure website, and enters the online banking facility using the customer number and credentials previously set

up. Each financial institution can determine the types of financial transactions which a customer may transact through online banking, but usually includes obtaining account balances, a list of recent transactions, electronic bill payments, financing loans and funds transfers between a customer's or another's accounts. Most banks set limits on the amounts that may be transacted, and other restrictions. Most banks also enable customers to download copies of bank statements, which can be printed at the customer's premises, though some banks charge a fee for mailing hard copies of bank statements. Some banks also enable customers to download transactions directly into the customer's accounting software. The facility may also enable the customer to order a cheque book, statements, report loss of credit cards, stop payment on a cheque, advise change of address and other routine actions.

Features of Online Banking

Online banking facilities typically have many features and capabilities in common, but also have some that are application specific. The common features fall broadly into following categories:

- A bank customer can perform non-transactional tasks through online banking, including:
  - Viewing account balances
  - Viewing recent transactions
  - Downloading bank statements, for example in PDF format
  - Viewing images of paid cheques
  - Ordering cheque books
  - Download periodic account statements
  - Downloading applications for M-Banking, E-Banking, etc.
- Bank customers can transact banking tasks through online banking, including:
  - Funds transfers between the customer's linked accounts
  - Paying third parties, including bill payments and third party fund transfers
  - Investment purchase or sale
  - Loan applications and transactions, such as repayments of enrolments
  - Credit card applications
  - Register utility billers and make bill payments
- Financial institution administration
- Management of multiple users having varying levels of authority
- Transaction approval process

NOTES

Some financial institutions offer special Internet banking services, for example:

- Personal financial management support, such as importing data into personal accounting software. Some online banking platforms support account aggregation to allow the customers to monitor all of their accounts in one place whether they are with their main bank or with other institutions.

Security

Security of a customer's financial information is very important, without which online banking could not operate. Similarly the reputational risks to banks themselves are important. Financial institutions have set up various security processes to reduce the risk of unauthorized online access to a customer's records, but there is no consistency to the various approaches adopted. The use of a secure website has been almost universally embraced.

Though single password authentication is still in use, it by itself is not considered secure enough for online banking in some countries. Basically there are following two different security methods in use for online banking:

The PIN/TAN System: The PIN/TAN system where the PIN represents a password, used for the login and TANs representing one-time passwords to authenticate transactions. TANs can be distributed in different ways, the most popular one is to send a list of TANs to the online banking user by postal letter. Another way of using TANs is to generate them by need using a security token. These token generated TANs depend on the time and a unique secret, stored in the security token (Two-Factor Authentication or 2FA). Usually online banking with PIN/TAN is done via a web browser using SSL secured connections, so that there is no additional encryption needed.

Signature Based Online Banking: Signature based online banking where all transactions are signed and encrypted digitally. The Keys for the signature generation and encryption can be stored on smartcards or any memory medium, depending on the concrete implementation.

13.3 AUTOMATED TELLER MACHINE OR ATM

An Automated Teller Machine (ATM) is an electronic telecommunications device that enables customers of financial institutions to perform financial transactions, such as cash withdrawals, deposits, funds transfers, or account information inquiries, at any time and without the need for direct interaction with bank staff.

ATMs are known by a variety of names, including Automatic Teller Machine (ATM) in the United States, sometimes redundantly as 'ATM machine'. In Canada, the term Automated Banking Machine (ABM) is also used, although ATM is also very commonly used in Canada, with many Canadian organizations using ATM over ABM. In British English, the terms cashpoint, cash machine and hole in the

wall are most widely used. Many ATMs have a sign above them indicating the name of the bank or organisation that owns the ATM, and possibly including the networks to which it can connect.

NOTES

Using an ATM, customers can access their bank deposit or credit accounts in order to make a variety of financial transactions, most notably cash withdrawals and balance checking, as well as transferring credit to and from mobile phones. ATMs can also be used to withdraw cash in a foreign country. If the currency being withdrawn from the ATM is different from that in which the bank account is denominated, the money will be converted at the financial institution's exchange rate. Customers are typically identified by inserting a plastic ATM card or some other acceptable payment card into the ATM, with authentication being by the customer entering a Personal Identification Number (PIN), which must match the PIN stored in the chip on the card, if the card is so equipped, or in the issuing financial institution's database.

According to the ATM Industry Association (ATMIA), as of 2015, there were close to 3.5 million ATMs installed worldwide. However, the use of ATMs is gradually declining with the increase in cashless payment systems.

ATMs can be placed at any location but are most often placed near or inside banks, shopping centers/malls, airports, railway stations, metro stations, grocery stores, petrol/gas stations, restaurants, and other locations. ATMs are also found on cruise ships and on some US Navy ships, where sailors can draw out their pay. ATMs may be ON- and OFF-premises. ON-premises ATMs are typically more advanced, multi-function machines that complement a bank branch's capabilities, and are thus more expensive. OFF-premises machines are deployed by financial institutions and Independent Sales Organisations (ISOs) where there is a simple need for cash, so they are generally cheaper single function devices.

In recent times, countries like India and some countries in Africa are installing ATMs in rural areas, which are solar powered.

Financial Networks

Most ATMs are connected to Interbank Networks, enabling people to withdraw and deposit money from machines not belonging to the bank where they have their accounts or in the countries where their accounts are held (enabling cash withdrawals in local currency). ATMs rely on authorization of a financial transaction by the card issuer or other authorizing institution on a communications network. This is often performed through an ISO 8583 messaging system.

Many banks charge ATM usage fees. In some cases, these fees are charged solely to users who are not customers of the bank that operates the ATM; in other cases, they apply to all users.

ATMs typically connect directly to their host or ATM Controller on either ADSL or dial-up modem over a telephone line or directly on a leased line. Leased

NOTES

lines are preferable to Plain Old Telephone Service (POTS) lines because they require less time to establish a connection. Less-trafficked machines will usually rely on a dial-up modem on a POTS line rather than using a leased line, since a leased line may be comparatively more expensive to operate compared to a POTS line. That dilemma may be solved as high-speed Internet VPN connections become more ubiquitous. Common lower-level layer communication protocols used by ATMs to communicate back to the bank include SNA over SDLC, TC500 over Async, X.25, and TCP/IP over Ethernet.

In addition to methods employed for transaction security and secrecy, all communications traffic between the ATM and the Transaction Processor may also be encrypted using methods, such as SSL (Secure Socket Layer).

13.4 ELECTRONIC STOCK EXCHANGE AND ELECTRONIC TRADING

In finance, an electronic trading platform also known as an online trading platform, is a computer software program that can be used to place orders for financial products over a network with a financial intermediary. Various financial products can be traded by the trading platform, over a communication network with a financial intermediary or directly between the participants or members of the trading platform. This includes products, such as stocks, bonds, currencies, commodities, derivatives and others, with a financial intermediary, such as brokers, market makers, Investment banks or stock exchanges. Such platforms allow electronic trading to be carried out by users from any location and are in contrast to traditional floor trading using open outcry and telephone based trading. Sometimes the term trading platform is also used in reference to the trading software alone.

Electronic trading platforms typically stream live market prices on which users can trade and may provide additional trading tools, such as charting packages, news feeds and account management functions. Some platforms have been specifically designed to allow individuals to gain access to financial markets that could formerly only be accessed by specialist trading firms. They may also be designed to automatically trade specific strategies based on technical analysis or to do high-frequency trading.

Electronic trading platforms are usually mobile-friendly and available for Windows, iOS and Android.

A stock exchange, securities exchange, or bourse is a facility where stockbrokers and traders can buy and sell securities, such as shares of stock and bonds and other financial instruments. Stock exchanges may also provide facilities for the issue and redemption of such securities and instruments and capital events including the payment of income and dividends. Securities traded on a stock exchange include stock issued by listed companies, unit trusts, derivatives, pooled investment products and bonds. Stock exchanges often function as ‘Continuous

NOTES

Auction' markets with buyers and sellers consummating transactions at a central location, such as the floor of the exchange or by using an electronic trading platform.

To be able to trade a security on a certain stock exchange, the security must be listed there. Usually, there is a central location at least for record keeping, but trade is increasingly less linked to a physical place, as modern markets use electronic communication networks, which give them advantages of increased speed and reduced cost of transactions. Trade on an exchange is restricted to brokers who are members of the exchange. In recent years, various other trading venues, such as electronic communication networks, alternative trading systems and 'Dark Pools' have taken much of the trading activity away from traditional stock exchanges.

Initial public offerings of stocks and bonds to investors is done in the primary market and subsequent trading is done in the secondary market. A stock exchange is often the most important component of a stock market. Supply and demand in stock markets are driven by various factors that, as in all free markets, affect the price of stocks.

There is usually no obligation for stock to be issued through the stock exchange itself, nor must stock be subsequently traded on an exchange. Such trading may be off exchange or over-the-counter. This is the usual way that derivatives and bonds are traded. Increasingly, stock exchanges are part of a global securities market. Stock exchanges also serve an economic function in providing liquidity to shareholders in providing an efficient means of disposing of shares.

An Electronic Communication Network (ECN) is a type of computerized forum or network that facilitates the trading of financial products outside traditional stock exchanges. An ECN is generally an electronic system that widely disseminates orders entered by market makers to third parties and permits the orders to be executed against in whole or in part. The primary products that are traded on ECNs are stocks and currencies. ECNs are generally passive computer-driven networks that internally match limit orders and charge a very small per share transaction fee.

To trade with an ECN, one must be a subscriber or have an account with a broker that provides direct access trading. ECN subscribers can enter orders into the ECN via a custom computer terminal or network protocols. The ECN will then match contra-side orders, i.e., a sell-order is 'Contra-Side' to a buy-order with the same price and share count for execution. The ECN will post unmatched orders on the system for other subscribers to view. Generally, the buyer and seller are anonymous, with the trade execution reports listing the ECN as the party.

Some ECN brokers may offer additional features to subscribers, such as negotiation, reserve size, and pegging, and may have access to the entire ECN book that real-time market data regarding depth of trading interest.

ECNs are generally facilitated by electronic negotiation, a type of communication between agents that allows cooperative and competitive sharing of information to determine a proper price.

13.5 DATA SHARING AND DISSEMINATION

NOTES

Data sharing is the practice of making data used available to other persons on the Internet. Many funding agencies, institutions, and publication venues have policies regarding data sharing because transparency and openness are considered by many to be part of the scientific method.

In order to encourage data sharing and prevent the loss or corruption of data, various rules and policies are established on data archiving. Access to publicly archived data is a recent development in the history of Internet technology made possible by technological advances in communications and information technology. To take full advantage of modern rapid communication may require consensual agreement on the criteria underlying mutual recognition of respective contributions.

Data sharing may also indicate the sharing of personal information on a social media platform.

Data dissemination is the distribution or transmitting of statistical, or other, data to end users. There are many ways organisations can release data to the public, i.e., electronic format, CD-ROM and paper publications, such as PDF files based on aggregated data. The most popular dissemination method today is the 'non-proprietary' open systems using Internet Protocols (IP). They are used in data dissemination through various communication infrastructures across any set of interconnected networks. Data is made available in common open formats.

Some organisations choose to disseminate data using 'Proprietary' databases in order to protect their sovereignty and copyright of the data. Proprietary data dissemination requires a specific piece of software in order for end users to view the data. The data will not open in common open formats. The data is first converted into the proprietary data format, and specifically designed software is provided by the organisation to users.

13.6 ELECTRONIC TRANSACTION AND DOCUMENT DELIVERY

An electronic transaction is the sale or purchase of goods or services, whether between businesses, households, individuals, governments, and other public or private organisations, conducted over computer-mediated networks. The goods and services are ordered over those networks, but the payment and the ultimate delivery of the good or service may be conducted ON - LINE or OFF - LINE.

Secure Electronic Transaction (SET) is a communications protocol standard for securing credit card transactions over networks, specifically, the Internet. SET was not itself a payment system, but rather a set of security protocols and formats that enabled users to employ the existing credit card payment infrastructure on an open network in a secure fashion. However, it failed to gain attraction in the market. Visa now promotes the 3-D Secure scheme.

NOTES

Moreover, the Secure Electronic Transaction (SET) is a system for ensuring the security of financial transactions on the Internet. It was supported initially by MasterCard, Visa, Microsoft, Netscape, and others. With SET, a user is given an electronic wallet (digital certificate) and a transaction is conducted and verified using a combination of digital certificates and digital signatures among the purchaser, a merchant, and the purchaser's bank in a way that ensures privacy and confidentiality.

Electronic Data Interchange (EDI) is the concept of businesses electronically communicating information that was traditionally communicated on paper, such as purchase orders and invoices. Technical standards for EDI exist to facilitate parties transacting such instruments without having to make special arrangements.

13.6.1 Authentication and Validation Transaction Processing

Electronic authentication is the process of establishing confidence in user identities electronically presented to an information system. Digital authentication or e-authentication may be used synonymously when referring to the authentication process that confirms or certifies a person's identity and works. When used in conjunction with an electronic signature, it can provide evidence of whether data received has been tampered with after being signed by its original sender. Electronic authentication can reduce the risk of fraud and identity theft by verifying that a person is who they say they are when performing transactions online.

There are various e-authentication methods that can be used to authenticate a user's identify ranging from a password to higher levels of security that utilize Multi-Factor Authentication (MFA). Depending on the level of security used, the user might need to prove his or her identity through the use of security tokens, challenge questions or being in possession of a certificate from a third-party certificate authority that attests to their identity.

The American National Institute of Standards and Technology (NIST) has developed a generic electronic authentication model that provides a basic framework on how the authentication process is accomplished regardless of jurisdiction or geographic region. According to this model, the enrollment process begins with an individual applying to a Credential Service Provider (CSP). The CSP will need to prove the applicant's identity before proceeding with the transaction. Once the applicant's identity has been confirmed by the CSP, he or she receives the status of 'Subscriber', is given an authenticator, such as a token and a credential, which may be in the form of a username.

The CSP is responsible for managing the credential along with the subscriber's enrollment data for the life of the credential. The subscriber will be tasked with maintaining the authenticators. An example of this is when a user normally uses a specific computer to do their online banking. If he or she attempts to access their bank account from another computer, the authenticator will not be present. In order to gain access, the subscriber would need to verify their identity to the CSP, which might be in the form of answering a challenge question successfully before being given access.

NOTES

Following are three generally accepted factors that are used to establish a digital identity for electronic authentication, including:

- Knowledge factor, which is something that the user knows, such as a password, answers to challenge questions, ID numbers or a PIN.
- Possession factor, which is something that the user has, such as mobile phone, PC or token.
- Biometric factor, which is something that the user is, such as his or her fingerprints, eye scan or voice pattern.

Out of the above mentioned three factors, the biometric factor is the most convenient and convincing to prove an individual's identity; but it is the most expensive to implement. Each factor has its weaknesses, so reliable and strong authentication depends on combining two or more factors. This is known as Multi-Factor Authentication (MFA). MFA can still be vulnerable to attacks, including man-in-the-middle attacks and Trojan attacks.

Check Your Progress

1. What is online banking?
2. Explain about the Automated Teller Machine (ATM).
3. What is an electronic trading platform?
4. What is data sharing and data dissemination?
5. Explain the terms electronic transaction and Secure Electronic Transaction (SET).

13.7 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. Online Banking, also known as Internet Banking or Web Banking, is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website. The online banking system will typically connect to or be part of the core banking system operated by a bank and is in contrast to branch banking which was the traditional way customers accessed banking services.
2. An Automated Teller Machine (ATM) is an electronic telecommunications device that enables customers of financial institutions to perform financial transactions, such as cash withdrawals, deposits, funds transfers, or account information inquiries, at any time and without the need for direct interaction with bank staff.

   Using an ATM, customers can access their bank deposit or credit accounts in order to make a variety of financial transactions, most notably cash

NOTES

withdrawals and balance checking, as well as transferring credit to and from mobile phones. ATMs can also be used to withdraw cash in a foreign country. If the currency being withdrawn from the ATM is different from that in which the bank account is denominated, the money will be converted at the financial institution's exchange rate. Customers are typically identified by inserting a plastic ATM card or some other acceptable payment card into the ATM, with authentication being by the customer entering a Personal Identification Number (PIN), which must match the PIN stored in the chip on the card, if the card is so equipped, or in the issuing financial institution's database.

3. In finance, an electronic trading platform also known as an online trading platform, is a computer software program that can be used to place orders for financial products over a network with a financial intermediary. Various financial products can be traded by the trading platform, over a communication network with a financial intermediary or directly between the participants or members of the trading platform. This includes products, such as stocks, bonds, currencies, commodities, derivatives and others, with a financial intermediary, such as brokers, market makers, Investment banks or stock exchanges.

   Electronic trading platforms are usually mobile-friendly and available for Windows, iOS and Android.

4. Data sharing is the practice of making data used available to other persons on the Internet. Many funding agencies, institutions, and publication venues have policies regarding data sharing because transparency and openness are considered by many to be part of the scientific method.

   Data dissemination is the distribution or transmitting of statistical, or other, data to end users. There are many ways organisations can release data to the public, i.e., electronic format, CD-ROM and paper publications, such as PDF files based on aggregated data. The most popular dissemination method today is the 'non-proprietary' open systems using Internet Protocols (IP). They are used in data dissemination through various communication infrastructures across any set of interconnected networks. Data is made available in common open formats.

5. An electronic transaction is the sale or purchase of goods or services, whether between businesses, households, individuals, governments, and other public or private organisations, conducted over computer-mediated networks. The goods and services are ordered over those networks, but the payment and the ultimate delivery of the good or service may be conducted ON - LINE or OFF - LINE.

   Secure Electronic Transaction (SET) is a communications protocol standard for securing credit card transactions over networks, specifically, the Internet. SET was not itself a payment system, but rather a set of security protocols

and formats that enabled users to employ the existing credit card payment infrastructure on an open network in a secure fashion.

NOTES

13.8 SUMMARY

- Online Banking, also known as Internet Banking or Web Banking, is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial institution's website.
- The online banking system will typically connect to or be part of the core banking system operated by a bank and is in contrast to branch banking which was the traditional way customers accessed banking services.
- Some banks operate as a 'Direct Bank' or 'Virtual Bank', where they rely completely on Internet banking.
- Internet banking software provides personal and corporate banking services offering features, such as viewing account balances, obtaining statements, checking recent transactions, transferring money between accounts, and making payments.
- To access a financial institution's online banking facility, a customer with Internet access will need to register with the institution for the service, and set up a password and other credentials for customer verification.
- The credentials for online banking is normally not the same as for telephone or mobile banking. Customer numbers are normally not the same as account numbers, because a number of customer accounts can be linked to the one customer number.
- Security of a customer's financial information is very important, without which online banking could not operate. Similarly the reputational risks to banks themselves are important.
- The PIN/TAN system where the PIN represents a password, used for the login and TANs representing one-time passwords to authenticate transactions.
- Usually online banking with PIN/TAN is done via a web browser using SSL secured connections, so that there is no additional encryption needed.
- Signature based online banking where all transactions are signed and encrypted digitally. The Keys for the signature generation and encryption can be stored on smartcards or any memory medium, depending on the concrete implementation.
- An Automated Teller Machine (ATM) is an electronic telecommunications device that enables customers of financial institutions to perform financial transactions, such as cash withdrawals, deposits, funds transfers, or account

information inquiries, at any time and without the need for direct interaction with bank staff.

NOTES

- Using an ATM, customers can access their bank deposit or credit accounts in order to make a variety of financial transactions, most notably cash withdrawals and balance checking, as well as transferring credit to and from mobile phones.
- ATMs can also be used to withdraw cash in a foreign country. If the currency being withdrawn from the ATM is different from that in which the bank account is denominated, the money will be converted at the financial institution's exchange rate.
- Customers are typically identified by inserting a plastic ATM card or some other acceptable payment card into the ATM, with authentication being by the customer entering a Personal Identification Number (PIN), which must match the PIN stored in the chip on the card, if the card is so equipped, or in the issuing financial institution's database.
- ATMs can be placed at any location but are most often placed near or inside banks, shopping centers/malls, airports, railway stations, metro stations, grocery stores, petrol/gas stations, restaurants, and other locations.
- Most ATMs are connected to Interbank Networks, enabling people to withdraw and deposit money from machines not belonging to the bank where they have their accounts or in the countries where their accounts are held (enabling cash withdrawals in local currency).
- ATMs rely on authorization of a financial transaction by the card issuer or other authorizing institution on a communications network. This is often performed through an ISO 8583 messaging system.
- In finance, an electronic trading platform also known as an online trading platform, is a computer software program that can be used to place orders for financial products over a network with a financial intermediary.
- Various financial products can be traded by the trading platform, over a communication network with a financial intermediary or directly between the participants or members of the trading platform. This includes products, such as stocks, bonds, currencies, commodities, derivatives and others, with a financial intermediary, such as brokers, market makers, Investment banks or stock exchanges.
- Electronic trading platforms typically stream live market prices on which users can trade and may provide additional trading tools, such as charting packages, news feeds and account management functions.
- Some platforms have been specifically designed to allow individuals to gain access to financial markets that could formerly only be accessed by specialist trading firms. They may also be designed to automatically trade specific strategies based on technical analysis or to do high-frequency trading.

NOTES

- Electronic trading platforms are usually mobile-friendly and available for Windows, iOS and Android.
- An Electronic Communication Network (ECN) is a type of computerized forum or network that facilitates the trading of financial products outside traditional stock exchanges.
- An ECN is generally an electronic system that widely disseminates orders entered by market makers to third parties and permits the orders to be executed against in whole or in part. The primary products that are traded on ECNs are stocks and currencies. ECNs are generally passive computer-driven networks that internally match limit orders and charge a very small per share transaction fee.
- Data sharing is the practice of making data used available to other persons on the Internet. Many funding agencies, institutions, and publication venues have policies regarding data sharing because transparency and openness are considered by many to be part of the scientific method.
- Data dissemination is the distribution or transmitting of statistical, or other, data to end users. There are many ways organisations can release data to the public, i.e., electronic format, CD-ROM and paper publications, such as PDF files based on aggregated data.
- The most popular dissemination method today is the 'non-proprietary' open systems using Internet Protocols (IP). They are used in data dissemination through various communication infrastructures across any set of interconnected networks. Data is made available in common open formats.
- An electronic transaction is the sale or purchase of goods or services, whether between businesses, households, individuals, governments, and other public or private organisations, conducted over computer-mediated networks. The goods and services are ordered over those networks, but the payment and the ultimate delivery of the good or service may be conducted ON - LINE or OFF - LINE.
- Secure Electronic Transaction (SET) is a communications protocol standard for securing credit card transactions over networks, specifically, the Internet.
- SET was not itself a payment system, but rather a set of security protocols and formats that enabled users to employ the existing credit card payment infrastructure on an open network in a secure fashion. However, it failed to gain attraction in the market. Visa now promotes the 3-D Secure scheme.
- Electronic authentication is the process of establishing confidence in user identities electronically presented to an information system.
- Digital authentication or e-authentication may be used synonymously when referring to the authentication process that confirms or certifies a person's identity and works. When used in conjunction with an electronic signature,

it can provide evidence of whether data received has been tampered with after being signed by its original sender.

- Electronic authentication can reduce the risk of fraud and identity theft by verifying that a person is who they say they are when performing transactions online.

NOTES

13.9 KEY WORDS

- **Online banking:** Online Banking, also known as Internet Banking or Web Banking, is an electronic payment system that enables customers of a bank or other financial institution to conduct a range of financial transactions through the financial (bank) institution's website.
- **PIN/TAN system:** The PIN/TAN system where the PIN represents a password, used for the login and TANs representing one-time passwords to authenticate transactions.
- **Signature based online banking:** Signature based online banking where all transactions are signed and encrypted digitally. The Keys for the signature generation and encryption can be stored on smartcards or any memory medium, depending on the concrete implementation.
- **Automated Teller Machine (ATM):** An Automated Teller Machine (ATM) is an electronic telecommunications device that enables customers of financial institutions to perform financial transactions, such as cash withdrawals, deposits, funds transfers, or account information inquiries, at any time and without the need for direct interaction with bank staff.
- **Electronic Communication Network (ECN):** An Electronic Communication Network (ECN) is a type of computerized forum or network that facilitates the trading of financial products outside traditional stock exchanges. An ECN is generally an electronic system that widely disseminates orders entered by market makers to third parties and permits the orders to be executed against in whole or in part.

13.10 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the term online banking.
2. What is an Automated Teller Machine (ATM)?
3. Explain the term Electronic Communication Network (ECN).
4. Differentiate between data sharing and data dissemination.
5. What is authentication and validation transaction processing?

Long-Answer Questions

NOTES

1. Briefly explain the process of online banking with reference to economic applications.
2. Explain the significance of Automated Teller Machine (ATM).
3. Elaborate on the terms electronic stock exchange and electronic trading giving appropriate examples.
4. Explain the concept of data sharing and dissemination giving appropriate examples.
5. Discuss about electronic transaction and document delivery. Also explain how the authentication and validation is performed for transaction processing. Support your answer giving appropriate examples.

13.11 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.

UNIT 14 E-BUSINESS

NOTES

Structure

14.0 INTRODUCTION

Online Business or e-business is any kind of business or commercial transaction that includes sharing information across the Internet. Commerce constitutes the exchange of products and services between businesses, groups and individuals and can be seen as one of the essential activities of any business. Electronic commerce focuses on the use of ICT to enable the external activities and relationships of the business with individuals, groups and other businesses, while e-business refers to business with help of the Internet. Electronic business differ from electronic commerce as it does not only deal with online transactions of selling and buying of a product and/or service but also enables to conduct business processes (inbound/outbound logistics, manufacturing & operations, marketing and sales, customer service) within the value chain through internal or external networks. The term 'E-Business' was coined by IBM's marketing and Internet team in 1996.

NOTES

Electronic business can take place between a very large numbers of market participants; it can be between business and consumer, private individuals, public administrations or any other organizations such as NGOs. These various market participants can be divided into three main groups, as Business (B), Consumer (C) and Administration (A). All of them can be either buyer or service provider within the market. There are nine possible combinations for electronic business relationships. B2C and B2B belong to e-commerce, while A2B and A2A belong to e-government sector that is also a part of electronic business.

A final way to secure information online would be to use a digital signature. If a document has a digital signature on it, no one else is able to edit the information without being detected. That way if it is edited, it may be adjusted for reliability after the fact. In order to use a digital signature, one must use a combination of cryptography and a message digest. A message digest is used to give the document a unique value. That value is then encrypted with the sender's private key

In this unit, you will study about the e-business: electronic trading and marketing; on-line shopping and malls, B2B, B2C, models, document and transaction security and digital signature; integrated transaction on mobile platforms, and e-commerce applications in India.

14.1 OBJECTIVES

After going through this unit, you will be able to:

- Understand the concept of e-business, electronic trading and marketing
- Explain about online shopping and malls
- Define the B2B and B2C models
- Elaborate on document and transaction security and digital signature
- Understand the process of integrated transaction on mobile platforms
- Discuss the e-commerce applications in India

14.2 E-BUSINESS

The following are the requirements of e-business:

- Improved customer service
- Origin of new business opportunities
- Enhanced speed and accuracy of a product
- Product cost saving

- **Improved Customer Services:** These days, consumers want better service. Therefore, e-business services offer a means of communication

between the consumer and the company. The consumer can even make online complaints to a company. Most websites provide a different e-mail id where complaints can be mailed. Customer-oriented organizations take complaints very seriously. Not only are the grievances given a good hearing. Actions is taken almost immediately. It is possible for satisfied as well as dissatisfied customers to express their opinions and also make suggestions. The bonding between the company and the customer strengthens.

- **Origin of New Business Opportunity:** Bigger network between consumers and companies can lead to new business opportunities. For example, a business may find infinite possibilities to develop and increase its consumer base. A company offering gardening-related products may also think of venturing into delivery of bouquets, cakes and gifts on request, for a price. Companies offering toys for toddlers could also come up with a forum where parents can interact with paediatricians or child psychologists to clarify doubts.
- **Enhanced Speed and Accuracy of a Product:** The usage of e-business services reduces human errors and other problems like a duplication of proceedings. This perfection in speed and accuracy, plus easy access to documents and information affect the increase in production. A customer care executive may often forget to enter necessary details of a transaction. She may call have to call up and bother the customer repeatedly to get some information. In case of online interaction, the customer will be filling in his own details. Reconfirmation may not be necessary at all. Wastage of time and money can be checked.
- **Product Cost Saving:** Despite the fact that you can reduce the cost of a product by the use of e-business services, it also reduces the errors and the cost of sending the information to partners.

Prerequisites of E-Business Procedure

More and more people are getting into e-business. This is natural because barriers to entry are hardly any. A lot of online businesses can be run from home. You do not really need degress and certificates to get started. However, you do need to give the venture some serious thought. A lot of work would be required, at least initially.

In order to conduct e-business, the main things you will require include the following:

1. A commercial website like www.futurebazaar.com
2. A product or service you want to sell through the respective websites
3. Shopping carts or purchase order forms
4. Current credit card account that will be accepted on e-payment

5. An online payment gateway, if you plan to process credit cards in real time, over the Internet
6. A Secure Socket Layer (SSL) to secure the gateway

NOTES

14.2.1 Functions of E-Business

E-business applications enable various business functions and transactions to be done electronically. Some of the functions are discussed as follows:

- **E-Advertising:** Advertising of information is currently the largest commercial activity on the Web. For example:
  - (i) A company's website contains its profile and all the information on its products and services.
  - (ii) It displays banners that can be clicked.
  - (iii) E-business portals like www.yahoo.com.
  - (iv) Newsgroups also provide publicity.
- **E-Catalogues:** Web pages that offer information on products or services that a company offers are available on an e-catalogue. An e-catalogue provides information on:
  - (i) Packaging
  - (ii) Product attributes and characteristics
  - (iii) Availability
  - (iv) Payment modes
  - (v) Cost, etc.
- **E-Publishing:** This sector was among the first few to spend on this novel technology especially on the Internet. E-publishing has led to several successful e-commerce endeavours, such as an independent publication through the Internet and electronic newspapers.

  Online publications offer services, such as:
  - (i) Online reading/browsing
  - (ii) Online search
  - (iii) Customized information services
- **E-Banking:** This facility offers remote banking electronically. Electronic banking is also referred to as online banking, cyber banking, home banking or virtual banking. It enables Web users to make online purchases and pay for the same using an online-banking facility. It is cost-effective, simple and available 24 hours. Customers have access to several services, such as:
  - (i) Bill payment
  - (ii) Electronic cheque writing

(iii) Record keeping

(iv) Tracking of bank account, credit cards

14.2.2 Shopping Services

NOTES

Shopping services are of many types. A few of them are discussed as follows:

1. Services provided by independent businesses who send their representatives to the stores to do comparison shopping for specific products. A shopping service is hired on contract to compare competitive prices or prices for the same item in competitive stores, depending on the request, and the needs of the client.
2. Shopping services that are offered to cable television subscribers where consumers can buy products (usually at a discount) that are displayed on a special shopping services channel.
3. Shopping services are offered to subscribers of personal information services for home computer use. For example, a company provides online information to subscribers. Among the many services offered by this company is one called products guides, from which consumers can shop and select purchases right from their own computer terminals.

14.2.3 Information Services

Information service is also known as information systems. For several organizations, information systems or information services are accountable for IT and Management Information Systems. Different types of decisions are supported by information systems at various levels of the organizational hierarchy. Key information systems include information management software and structural databases. They include the following:

- Enterprise Collaboration System (ECS)
- Transaction Process System (TPS)
- Decision Support System (DSS)
- Executive Support System (ESS)
- Management Information System (MIS)

14.2.4 E-Business Framework Architecture

E-business applications are built on the existing infrastructure for online communication, network and connection software which frames the nascent information superhighway.

E-Commerce Applications

E-business can be applied in:

1. Supply Chain Management

NOTES

2. Online Banking
3. Procurement and Purchasing
4. Online Marketing and Advertisement
5. Home Shopping

Figure 14.1 shows several e-business applications, including interorganizational and customer-oriented examples. Without each of the building blocks in the infrastructure, none of these uses will be possible.

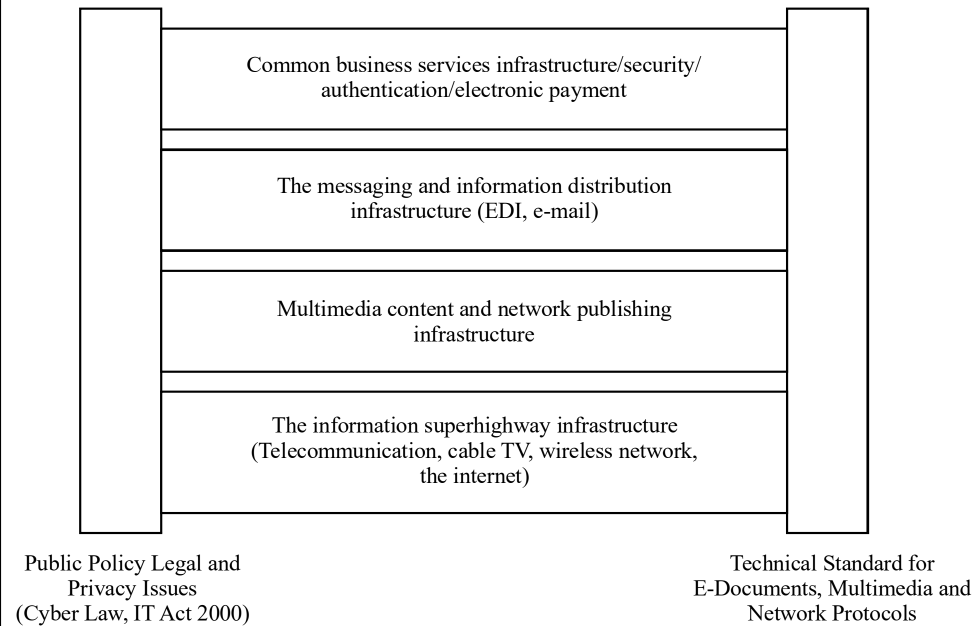


***Fig. 14.1** E-Business Applications*

Client–Server Architecture in E-Business

The client–server model is followed by the applications of e-business (Refer Figure 14.2). The model lets the client work together with the server with the help of a request–reply sequence governed by message passing. The server handles application tasks, storage and security and gives scalability to increase clients.

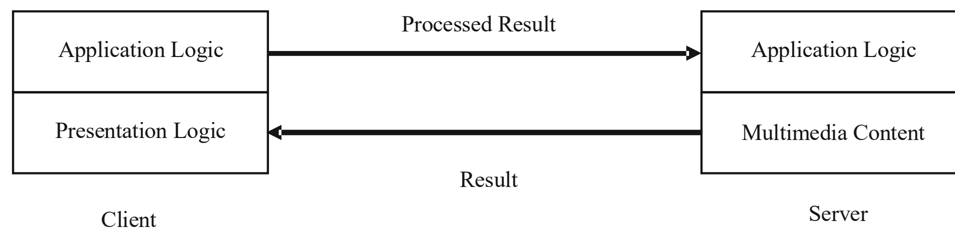


***Fig. 14.2** Client–Server Architecture*

Building Blocks of E-Business

Enterprise applications, insights, functions and IT infrastructure form the building blocks of e-business.

NOTES

Enterprise application software refers to software that facilitates the performance of business functions such as scheduling (of production/manufacturing processes, etc.), accounting, management of customer databases/information, management of bank account/dealings, etc. It is common to see such software hosted on servers. It is capable of serving multiple enterprises concurrently, over a computer network. This gives enterprise software an edge over the commonly used single-user applications which can only perform on a user's PC. These single-user applications can only serve one user at a time. Enterprise software provides solutions to problems that concern the enterprise as a whole and not individual departments. Only large enterprises can afford to build such enterprise software. This software becomes the pillar of the IT systems on which the entire enterprise functions and communicates. Among other things, there are enterprise applications for various functions such as:

- Finance
- HRM
- Customer Relationship Management or CRM
- Supply Chain Management or SCM
- Product Life Cycle or PLC Management
- Enterprise Commerce Management
- Enterprise Resource Planning or ERP

All these processes, such as SCM, CRM and ERP need to be integrated properly for e-business to be successful. The IT infrastructure required for the same may include the following:

- Application Servers
- Web Services
- Wireless Technology
- Database Software
- XML
- Storage Systems
- Server Platform

Business intelligence would also be required for Web analysis, for managing knowledge and content and for mining data, if required, depending on the nature of e-business.

For today's growing e-businesses, it is very important to consider the issues of security of information and the reliability and scalability of the systems as well. This is taken care of by integration.

NOTES

The two indispensable pillars that support all e-business applications and infrastructure are as follows:

(i) Public policy to govern universal access to privacy and information pricing.

(ii) User interface and transport in the interest of compatibilities across the entire network to dictate the nature of information publishing. Information should be accessible by any device, which the consumer chooses, and should be supported by any type of operating system.

14.3 ONLINE SHOPPING AND MALLS

Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the Internet using a web browser. Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine, which displays the same product's availability and pricing at different e-retailers. As of 2020, customers can shop online using a range of different computers and devices, including desktop computers, laptops, tablet computers, smartphones, and smart speakers.

An online shop evokes the physical analogy of buying products or services at a regular 'Bricks-and-Mortar' retailer or shopping center; the process is called Business-To-Consumer (B2C) online shopping. When an online store is set up to enable businesses to buy from another businesses, the process is called Business-To-Business (B2B) online shopping. A typical online store enables the customer to browse the firm's range of products and services, view photos or images of the products, along with information about the product specifications, features and prices.

In the twenty first century, online shopping has become very popular, especially with the lifestyles of business people who are always busy and are looking for a convenient way to shop.

In 1991 the World Wide Web (WWW) opened for commercial use. In 1994 other advances took place, such as online banking and the opening of an online pizza shop by Pizza Hut. During that same year, Netscape introduced Secure Sockets Layer (SSL) encryption of data transferred online, which has become essential for secure online shopping. In 1995, Amazon launched its online shopping site, and in 1996, eBay appeared.

A good online store is a website that is easy to navigate and browse for possible purchases. It has a product catalog that customers can use to browse the

search criteria (usually a type, price, material, age, etc.), and information about the products, the sellers, and the service center. Online stores may also discuss business conditions and a Complaints Procedure.

NOTES

If a shopper finds a product to purchase, clicking 'Send' will add the item to the shopping cart. The shopping cart collects all items to be purchased. Once a shopper is satisfied with their selections, the shopper makes a binding order and payment using a credit card or other financial arrangement. Therefore, implementing usability testing is highly important for an online store to avoid any error and improve the overall performances of the online store.

Online stores usually enable shoppers to use 'Search' features to find specific models, brands or items. Online customers must have access to the Internet and a valid method of payment in order to complete a transaction, such as a credit card, an Interac-enabled debit card, or a service, such as PayPal. For physical products, for example paperback books or clothes, the e-tailer ships the products to the customer; for digital products, such as digital audio files of songs or software, the e-tailer usually sends the file to the customer over the Internet. The largest of these online retailing corporations are Alibaba, Amazon.com, and eBay.

One of the earliest forms of trade conducted online was IBM's OnLine Transaction Processing (OLTP) developed in the 1960s and it allowed the processing of financial transactions in real-time. The emergence of online shopping as we know today developed with the emergence of the Internet. Initially, this platform only functioned as an advertising tool for companies, providing information about its products. It quickly moved on from this simple utility to actual online shopping transaction due to the development of interactive Web pages and secure transmissions. Specifically, the growth of the Internet as a secure shopping channel has developed since 1994, with the first sales of Sting album "Ten Summoner's Tales". Chocolates, and flowers soon followed and were among the pioneering retail categories which fueled the growth of online shopping. Researchers found that having products that are appropriate for e-commerce was a key indicator of Internet success. Many of these products did well as they are generic products which shoppers did not need to touch and feel in order to buy.

Online customers must have access to the Internet and a valid method of payment in order to complete a transaction. Generally, higher levels of education and personal income correspond to more favorable perceptions of shopping online. Increased exposure to technology also increases the probability of developing favorable attitudes towards new shopping channels.

Customer Buying Behaviour in Digital Environment

The marketing around the digital environment, customer's buying behaviour may not be influenced and controlled by the brand and firm, when they make a buying decision that might concern the interactions with search engine, recommendations, online reviews and other information. With the digital devices environment, people

NOTES

are more likely to use their mobile phones, computers, tablets and other digital devices to gather information. In other words, the digital environment has a growing effect on consumer's mind and buying behaviour. In an online shopping environment, interactive decision may have an influence on aid customer decision-making. Each customer is becoming more interactive, and though online reviews customers can influence other potential buyers' behaviors.

Subsequently, risk and trust would also are two important factors affecting people's' behavior in digital environments. Customer consider to switch between e-channels, because they are mainly influence by the comparison with offline shopping, involving growth of security, financial and performance-risks In other words, a customer shopping online that they may receive more risk than people shopping in stores. There are three factors may influence people to do the buying decision, firstly, people cannot examine whether the product satisfy their needs and wants before they receive it. Secondly, customer may concern at after-sale services. Finally, customer may afraid that they cannot fully understand the language used in e-sales. Based on those factors customer perceive risk may as a significantly reason influence the online purchasing behaviour.

Online retailers has place much emphasis on customer trust aspect, trust is another way driving customer's behaviour in digital environment, which can depend on customer's attitude and expectation. Indeed, the company's products design or ideas cannot met customer's expectations. Customer's purchase intension based on rational expectations, and additionally impacts on emotional trust. Moreover, those expectations can be also establish on the product information and revision from others.

Product Selection

Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine. Once a particular product has been found on the website of the seller, most online retailers use shopping cart software to allow the consumer to accumulate multiple items and to adjust quantities, like filling a physical shopping cart or basket in a conventional store. A "Checkout" process follows (continuing the physical-store analogy) in which payment and delivery information is collected, if necessary. Some stores allow consumers to sign up for a permanent online account so that some or all of this information only needs to be entered once. The consumer often receives an e-mail confirmation once the transaction is complete.

Shopping Mall

A mall or shopping centre is a large building that is full of many smaller shops and stores. It is different from earlier markets or bazaars because most of the shops are not little booths or stalls in one big open area. Each store has its own space with walls. Most of their entrances face a central walking area inside the building.

People visit the stores in the mall to shop. Most malls have parking lots (places to park cars). Most malls also have roofs so people can shop inside. Most malls have a food court. Food courts are areas where people buy food. Some malls have movie theatres also.

A shopping mall or simply mall is a North American term for a large indoor shopping center, usually anchored by department stores. The term originally meaning a pedestrian promenade with shops along it, but in the late 1960s began to be used as a generic term for the large enclosed shopping centers becoming common at that time. Successful exceptions have added entertainment and experiential features, added big-box stores as anchored, or are specialized formats: power centers, lifestyle centers, factory outlet centers, and festival marketplaces.

The International Council of Shopping Centers (ICSCs) classifies two types of shopping centers as malls: Regional Malls and Superregional Malls.

Regional Mall: A regional mall is as per the International Council of Shopping Centers, in the United States, a shopping mall with 400,000 sq ft (37,000 m^2) to 800,000 sq ft (74,000 m^2) gross leasable area with at least two anchor stores.

Superregional Mall: A superregional mall is, per the International Council of Shopping Centers, in the US is a shopping mall with over 800,000 sq ft (74,000 m^2) of gross leasable area, three or more anchors, mass merchant, more variety, fashion apparel, and serves as the dominant shopping venue for the region (25 miles or 40 km) in which it is located.

In India, the Internet users are about 475 million as of July 2019, about 40% of the population. In India, cash on delivery is the most preferred payment method, accumulating 75% of the e-retail activities. Demand for international consumer products (including long-tail items) is growing faster than in-country supply from authorised distributors and e-commerce offerings. Long tail business strategy allows companies to realize significant profits by selling low volumes of hard-to-find items to many customers, instead of only selling large volumes of a reduced number of popular items.

In 2017, the largest e-commerce companies in India were Flipkart, Snapdeal and Amazon. In 2018, Amazon beat Flipkart and was recorded the biggest ecommerce in India in terms of revenue.

14.4 BUSINESS MODELS IN PRACTICE

There are three fundamental types of business models in practice. These are:

1. Business-to-Business (B2B)
2. Business-to-Consumer (B2C)
3. Consumer-to-Consumer (C2C)

Table 14.1 Summary of E-Commerce Model

Model	Description	Example of Website
Business-to-Consumer	Goods or services sold directly to consumers	pets.com, edirects.com, amazon.com, autobytel.com
Business-to-Business	Goods or services sold between business and other businesses	verticalnet.com, metalsite.com, shop2gether.com
Business-to-Government	Goods or services sold to government agencies	igov.com
Consumer-to-Consumer	Goods or services sold between consumers	ebay.com, inforocket.com
Consumer-to-Business	Consumers fix the cost of their goods or services for other consumer	priceline.com

NOTES

1. Business-To-Business Model

The Business-To-Business (B2B) model needs two or more business organizations that do business with each other. It entails commercial activity among companies through the Internet as a medium. At present, there are many types of e-businesses. The B2B e-business is of the following types:

(a) Supplier Oriented

In this type of B2B e-business, a supplier establishes the electronic market where a number of customers or buyers transact with suppliers. Generally, it is done by a supplier which has monopoly over products that it supplies.

(b) Buyer Oriented

In this type of B2B electronic commerce, big business organizations with high volume purchase capacity creates an e-business marketplace for purchases and gains by starting a site of their own. The online e-business marketplace is used by buyers for placing requests for quotations and carrying out the entire purchase process.

(c) Intermediary Oriented

In this type of B2B e-business, a third party establishes the e-business marketplace and attracts both buyers and sellers to interact with each other.

Application of B2B Model

Some of the applications of B2B model are, inventory management, channel management, distribution management, order fulfilment and delivery payment and payment management.

2. Business-To-Consumer Model

The business to consumer model clearly concentrates on individual buyers and is thus known as Business-To-Consumer (B2C) model. The B2C model offers consumers the capabilities to browse, select and merchandise online from a wider

variety of sellers and at better prices. The B2C e-business interaction is most appropriate for the following types of transactions:

(i) Easily transformable goods, i.e., products that are easily transformable into digital format, such as videos, software packages, music books, and so on

(ii) Highly rated branded items or items with return security

(iii) Items sold in packets that are not possible to open in physical stores

(iv) Items that follow standard specification

NOTES

The following steps summarizes the working of B2C:

(i) The customer identifies his/her need.

(ii) Then, the customer looks for the product or service that suit his/her needs.

(iii) The customer selects a vendor and negotiates a price.

(iv) The customer then receives the product or service.

(v) The customer makes the payment for the received product service.

(vi) The customer gets the services and warranty claims that are associated with the product.

3. Consumer-To-Consumer Model

In a Consumer-To-Consumer (C2C) model, consumers sell directly to other consumers via online classified advertisements and auctions or by selling personal services or expertise online. The C2C model involves the growing popularity of Peer-To-Peer (P2P) software that facilities the exchange of data directly between individuals over the Internet.

Table 14.2 *Summary of Business Models*

Model	Description	Example
B2B	One business sells products or services to other businesses	metalsite.com
B2C	Business sells products or services directly to consumers	amazon.com
C2C	Consumer sells directly to other consumers	ebay.com

14.4.1 Business Model: The Six Components

According to Henry Chesbrough and Richard S. Rosenbloom there are six components of a business model, namely :

(i) Value proposition

(ii) Market segment

(iii) Value chain structure

(iv) Revenue generation and margins

NOTES

(v) Position in the value network

(vi) Competitive strategy.

(i) ***Value Proposition*****:** It has three components as follows:

(a) It is an explanation of any problem faced by a customer.

(b) It is about the resolution of that problem

(c) It is the value of this resolution from the customer's point of view.

(ii) ***Market Segment*****:** Since diverse market segments have different requirements, it boils down to which group to target. At times, the full benefit of a new product development is realized only when a new market segment is focused.

(iii) ***Value Chain Structure*****:** The concept of value chain structure demonstrates the company's place and the value addition tasks done by it in the value chain. It also shows in what ways the company captures part of the value, which it has helped to add in the chain.

(iv) ***Creation of Revenue and Profits*****:** It means how income is created in the business, such as rental, sales, subscription, and so on. It also includes revenue made from the target profit margins and the cost structure.

(v) ***Place in the Value Network:*** It identifies competitors and sellers whose products, services or relationships create more demand for your product. It also looks for other effects in the business network that might be used to give more value to the customer.

(vi) ***Competitive Strategy*****:** It refers to the ways in which a business organization tries to expand a permanent competitive benefit and utilize it to advance the competitive situation of the company in the market.

Business Model vs Revenue Model

The term 'business model' is a broad term that explains things, such as the place of a business organization in the value chain, the choice of the customers, products and the cost of doing business. The business model plans the course through which the company would in fact make profit. It statis clearly the price it would charge the customers.

The old model corporations have been discredited as overmanaged, overcontrolled and overstructured but underlet. Thus, today's top managers should focus on their real managerial skills that would bring success to the firm in the future. Today, a brand latest business model has come up. Here many of the important tasks of the company are delegated to the different individual parts, but synergy results from the vigour of the employees and unrestricted flow of information.

Role of a Business Model

To profit from an innovation in a product or a service, a new firm or a start-up needs an appropriate business model so that it is able to exploit its innovation and be the market leader. Business models are needed to bring in new technology that will yield an economic value. As the old and familiar business models cannot be used for all new firms, new business models are planned. The importance of the business model cannot be denied, because in many cases the profitability of the innovation rests more on the business model itself than on the product or service provided by the innovation. In their paper, 'The Role of the Business Model in Capturing Value from Innovation', Henry Chesbrough and Richard S. Rosenbloom provide a crucial structure explaining the basics of a business model. As there is a complex inter-play of markets, products and the environment in which a business organization runs, it is very difficult to understand the organization's responsibility in its totality. While business experts are acquainted in their area, technical experts understand theirs. Figure 14.3 makes it clear how the business model serves to connect these two domains.



Fig. 14.3 *Role of a Business Model*

Many business subjects including finance, economics, entrepreneurship, marketing operations and strategy are used to finalize a business model. The business model itself is an essential determinant of the profits to be generated from an innovation. A below average innovation with a great business model could be more profitable than a good innovation with a below average business model.

Business Model vs Business Strategy

You have read earlier in this unit about the six components of business models by Chesbrough and Rosenbloom. They further strike a comparison between the concept of the business model to that of strategy, identifying the following three differences:

(i) ***Creating Value vs Capturing Value:*** The focus of the business model is on the creation of value. Though the business model only addresses how that value would be captured by the organization, strategy focusses on building a sustainable competitive advantage.

(ii) ***Business Value vs Shareholder Value:*** The business model helps in the conversion of innovation to yield economic value for the businesses, but it does not focus on delivering business value to the shareholders. For instance,

NOTES

though the business model does not consider the financing methods, nonetheless, it impacts shareholder value.

(iii) ***Assumed Knowledge Levels:*** Business model assumes a limited environmental knowledge, even though strategy is dependent on a more intricate examination that needs more conviction about the environment.

Advantages of Entrepreneurship

According to Chesbrough and Rosenbloom, a good business model like Xerox has a tendency to establish thrust, but the company remains constrained to its thriving model. At the same time, the coming up of new technologies forces business organizations to evolve new business models. This gives new companies or start-ups a freehand to make a choice or even develop a new business model themselves. Otherwise, in adddtion to the risk taken up in the technology and economic areas, an unproven business model increases the risk further. Business ventures, generally, are more prepared to acknowledge this risk.

On the other hand, many venture capitalists fancy themselves as investors in business models. Thus, when it becomes obvious that the previous model is not working, the venture capitalists often try and push for a change in the business model.

14.5 DOCUMENT AND TRANSACTION SECURITY AND DIGITAL SIGNATURE

14.5.1 Cryptography

Cryptography is derived from the Greek words *kryptos* (hidden, secret) and *gráphô* (I write). It is the practice and study of hiding information. Cryptography is today considered a branch of both mathematics and computer science, and is used extensively in information theory, computer security and engineering. Cryptography is used in applications which require security of data, such as in the case of ATM cards, computer passwords and electronic commerce.

Purpose of Cryptography

The science of writing in a secret code is called '**Crypto**'. It is supposed to have been first used as far back as 1900 BC by an Egyptian scribe. Cryptography is believed to have appeared soon after writing was invented and used in diplomatic exchanges and battle plans. With the development of computer communication, the need for security of communication media also rose. Quite understandably, then, cryptography began to be used to provide this security while communicating over any untrusted medium, particularly the Internet.

These security requirements include:

- Authentication: That is, giving proof of one's identity.
- Privacy/confidentiality: Making sure that no one other than the intended reader reads the message.
- Integrity: Providing assurance to the receiver that the message received by him is no different from the original one.
- Non-repudiation: A mechanism which will prove that the message was actually sent by the sender and no one else.

Thus, it is seen that cryptography serves a dual purpose: data is protected from being stolen or altered and users are authenticated. This is done in three ways: (a) Secret Key (or Symmetric) Cryptography, (b) Public-Key (or Asymmetric) Cryptography and (c) Hash Functions. The unencrypted data is referred to as *plaintext*. It is encrypted into *ciphertext*, and then decrypted into usable plaintext.

Encryption as the Basis for Data and Messaging Security

Encryption is a cryptography technology to scramble (encrypt) the data with a key so that no one can make sense of it while it is being transmitted. When the data reaches its destination, the information is unscrambled (decrypted) using the same or a different key.

The terms used commonly in a cryptography system are as follows:

- **Intruder:** An intruder is any person who does not have the authorization to access the network or the information.
- **Plaintext:** It is an intelligible message that needs to be converted into an unreadable message or encrypted message.
- **Ciphertext:** A message in an encrypted form.

Example:

	(Encrypted Form)		(Decrypted Form)	
Plain Text	**Algorithm**	**Cipher Text**	**Algorithm**	**Plain Text**
Goods	Next two letters	Iqqfu	Previous two letters	Goods
Sales	Previous one letter	rzkdr	Next one letter	Sales

Encryption is a method by which plaintext can be converted into ciphertext.
Decryption is a method by which a ciphertext can be converted into plaintext.
Algorithm: A cryptography algorithm is a mathematical function.
Key: It is a string of digits.

14.5.2 Encryption

Methods of Encryption

NOTES

There are three types of cryptography or methods of encryption:

- Secret Key or Private Key or Symmetric Key Cryptography
- Public Key or Asymmetric Key Cryptography
- Hash Function

1. Secret Key or Symmetic Key Cryptography

In this scheme, both the sender and the recipient possess the same key to encrypt and decrypt the data. Figure 14.4 shows how secret or private key cryptography works.

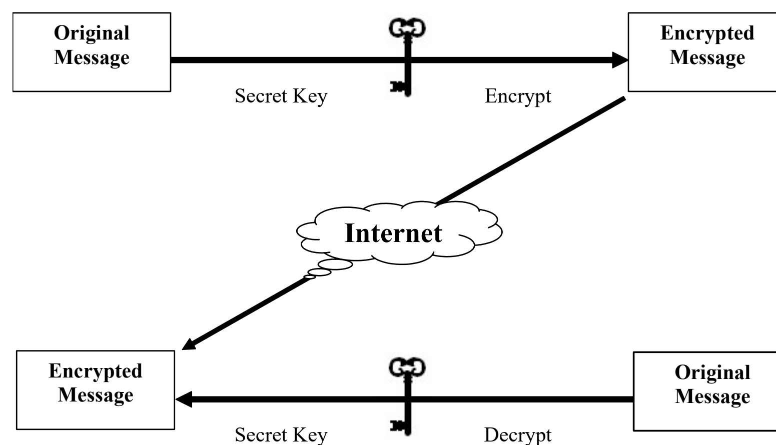


Fig. 14.4 *Schematic Diagram of Secret Key Cryptography*

Data Encryption Standard

Data Encryption Standard (DES) is an example of secret key cryptography. It was developed by IBM. DES is block cipher-based scheme which encrypts a 64-bit data block using a 56-bit key. The block is transformed in such a way that it involves sixteen iterations. This done by using the security key.

To take an example, suppose, A encrypts a message with a secret key and e-mails it to B, who on receiving it, checks the header to identify the sender. B then has to take the duplicate of the secret key to decrypt the message.

Drawbacks of Secret Key Cryptography:

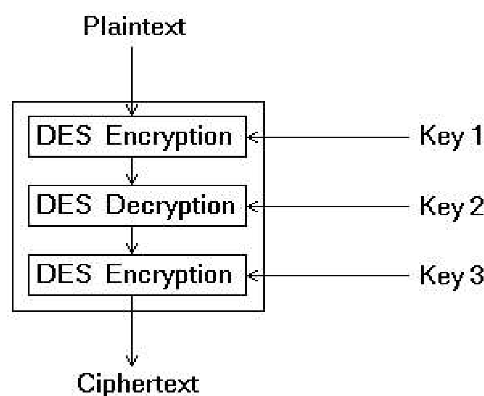
- Both parties must agree upon a shared secret key.
- If there are *n* correspondents, you have to keep track of *n* different secret keys. If the same key is used by more than one correspondent, the common key holders can read each other's mail.

- Symmetric encryption schemes are also subject to authenticity problems. Since both the sender and the recipient have the same secret key, the identity of originator or recipient cannot be proved. Both can encrypt or decrypt the message.

2. Triple Encryption

As discussed, the DES is a block cipher and employs shared secret encryption. But, nowadays DES is considered unsafe for various applications primarily due to the 56-bit key size which is too small. Triple DES is considered as an improved version to overcome many of the shortcomings of DES. The triple encryption technology is based on DES and is sometimes referred as Triple DES or 3DES. The event follows an Encrypt-Decrypt-Encrypt (EDE) sequence. Decrypt sequence is just the same encrypting operation with the keys reversed. It is based on the DES algorithm and can easily modify the existing software to use Triple DES. It has a longer key length that helps in eliminating many of the shortcut attacks used to reduce the amount of time it takes to break DES. Thus, Triple DES is considered as an exceptional and dependable option to fulfill the security requirements of highly sensitive information.

Triple DES mode of operation takes three 64-bit keys for an overall key length of 192 bits. In Private Key Encryption, the user can just type in the complete 192-bit (24 character) key rather than entering each of the three keys individually. The procedure for encryption is exactly the same as regular DES, but it is repeated three times. The data is encrypted with the first key, decrypted with the second key and finally encrypted again with the third key.



3. Public Key Cryptography

This scheme operates on a double key, called pair key, one of which is used to encrypt the message and the other is used to decrypt it. This can be viewed as two parts; one part of the key pair, the private key, is known only by the designated owner. The other part, the public key, is published widely but is still associated with the owner of the private key. Figure 14.5 shows how public key encryption works.

NOTES

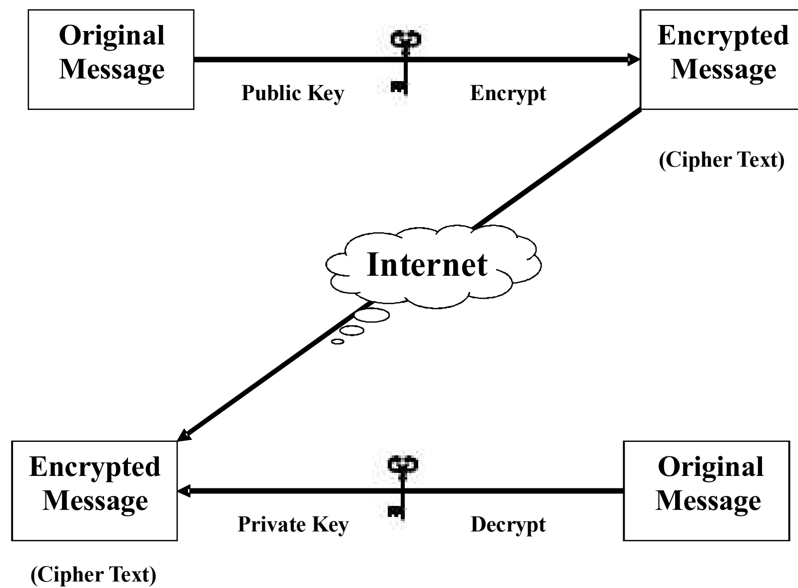


Fig. 14.5 *Schematic Diagram of Public Key Cryptography*

Encryption and Decryption

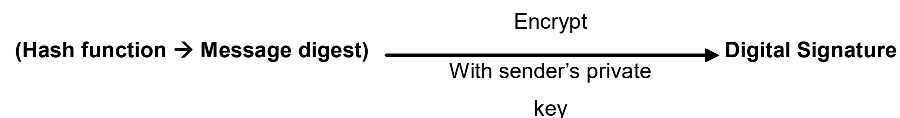
- Data encrypted with a public key can only be decrypted with a private key.
- Data encrypted with a private key can only be decrypted with a public key.

Advantages of Public Key Cryptography

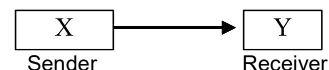
- ***Message confidentiality can be proved:*** The sender uses the recipient's public key to encrypt a message, so that only the private key holder can decrypt the message, and no one else.
- ***Authenticity of the message originator can be proved:*** The receiver uses his private key to encrypt a message, to which only the sender has access.
- ***Easy to distribute public key:*** The public key of the pair can be easily distributed.

Hash Function

Hash function is a formula that converts a message of a given length into a string of digits called a message digest. A mathematical transformation is used by the hash function to encrypt information such that it is irreversible. The encrypted ciphertext message cannot be decrypted back to plaintext.



How it works: X sends a message to Y



(a) The sender generates a message.

(b) A 'Message Digest' of the message is created using the hash function.

(c) The sender attaches the digital signature to the end of the message.

(d) The sender encrypts both message and signature with the receiver's public key.

(e) Using a private key, the entire message is encrypted by the receiver.

(f) The receiver calculates the message digest using the hash function.

The receiver uses the same hash function that the sender uses, and which has been agreed upon in advance. The main advantage of using the hash function for encryption is that even if an unauthorized person accesses X's public key, he will not be able to get to the hash function-generated key; thus making the digital signature authentic and secure.

NOTES

14.5.3 Digital Signature

Digital signatures are used for authenticating e-commerce business transactions. The authentications refer to legal, financial and other document-related issues. Digital signatures are just like handwritten signatures which determine authentications.

A digital signature consists of two parts:

(i) Signature in the document: signer authentication

(ii) Document authentication

(i) **Signer Authentication:** A signature should indicate who signed a document, message or record and should be difficult for another person to produce without authorization.

(ii) **Document Authentication:** A signature should identify what is signed so that:

- Sender can not remove the content of messages after signing it.
- The receiver cannot make any changes in the message.

Validity of Digital Signatures

Generally, a key expires after a certain period that could range from six months to a year. A signed document with an expired key is not acceptable. The contract is registered with a digital time stamping service at the time it is signed; the signature can be authenticated even after expiry of the key. If every party on the contract keeps a copy of the timestamp, all of them can prove that the contract was signed using valid keys. Actually, the timestamp can prove the validity of the contract even if one signatory's key gets compromised at any instant after the contract is signed.

NOTES

A Digital Time Stamping (DTS) service issues timestamps which associate a date and time with a digital document in a cryptographically strong way. The digital timestamp can be used at a later date to prove that an electronic document existed at the time stated on its timestamp.

Because keys are intended to be public and are widely distributed, anyone can easily create a private/public key pair and distribute the public key, claiming it belonged to someone else. One solution to this problem is a public-key certificate. A public-key certificate is a data structure, digitally signed by a Certifying Authority (CA).

Certificates Authority

Certificates authority is an organization or institution that issues digital certificate to companies and organizations that are accessible via the Internet. These certificates are issued for a certain period of time and are used as an assurance of the security of a website. It is also known as trusted third party. CAs form characteristics of many Public Key Infrastructure (PKI) schemes. There are many commercial CAs that charge for their services. There are also several providers issuing digital certificates to the public without any cost. Generally, institutions and governments have their own CAs.

Certificates authority issues digital certificates that consist of the identification details of the owner and his public key. The corresponding private key is in a similar manner not made available publicly, but kept as a secret by the end-user, who generates a key pair. The certificate also acts as evidence by the CA such that the public key contained in the certificate is related to the person, organization, server or other entities noted in the certificate. If the user believes in the Certificate Authority (CA) and is able to validate the CA's signature, then he can also validate the requirement of a certain public key that belongs to whoever is identified in the certificate.

Digital Certificate

A digital certificate serves as an electronic identity card that establishes the user's credentials when business deals are transacted across the Web. A digital certificate is defined as a method to electronically verify for authenticity. The digital certificate is just like an identity card, such as a driver's license. Digital certificate is issued by a number of certificate authorities; it is used to prove that a website, or a visitor to a website, is the entity or person they claim to be; An electronic credential issued by a certification authority to establish the identity of an organization when doing business on the Internet.

Contents of Digital Certificate

A digital certificate contains the following details:

- Certificate Holder's Name, organization and address.

- The name of certificate authority who has issued this certificate.
- Public key of the holders for cryptographic use.
- Time limit, these certificates are issued for durations of six months to a year.
- Digital certificate identification number.

NOTES

A digital certificate contains a public key that is used for encrypting messages and digital signatures. It also has the digital signature of the certificate authority. By this signature a recipient can verify that the certificate is genuine. Sometime-digital certificates conform to a standard, X.509. It can be kept in registries so that authenticating users can look up other users' public keys.

Non-Repudiation and Message Integrity

Digital identity is based on message integrity, non-repudiation and confidentiality. Integrity ensures that a message or transaction has not been tampered with. Non-repudiation ensures that the contents of the message sent are intact and provides evidence for the existence of a message or transaction. The user and the recipient cannot dispute the contents, once sent. The contents are protected as confidential which means that only authorized individuals or groups can access the contents of a message or transaction. In certain cases, these features are not necessary and hence are considered as luxury. However, there are scenarios where these features are most critical. For managing digital identity strategy, clarity of these features is very important.

Integrity: Integrity is the basic requirement of a highly dependable identity infrastructure. Identity systems serve the purpose of exchanging credentials as well as messages and transactions pertaining to attributes, provisioning of information and other data. Integrity builds a trust that the contents have not been tampered, which is important in this environment. To understand this better, let us take an example of a document that represents identity credentials. It is important to validate the authenticity of the credentials to be sure of their originality.

Non-Repudiation: Non-repudiation is the activity of presenting of tamper-proof evidence proving that a message was sent or received. Critical identity-related acts should be protected even though the messages or transactions can be disputed. For understanding this better, let us take the instance of two people, Nadia and Joe, who are exchanging messages. In one case, Nadia denies sending a message to Joe that he claims to have received. The ability to counter Nadia's denial is called Non-Repudiation of Origin (NRO). In the second scenario, Nadia claims to have sent Joe a message that he denies having received. Provision of evidence to counter Bob's claim is called Non-Repudiation of Receipt (NRR).

14.6 INTEGRATED TRANSACTION ON MOBILE PLATFORMS

NOTES

Since the development of mobile wireless technology in the early 1970s, the mobile wireless industry has come up with many new technologies that have revolutionized the way communication occurs the world over. The recent past has seen the mobel wireless technologies have seen four or five generations starting from 0G–4G. With the introduction of the cellular concept in the 1G brought about the mobile wireless communication. Since then the 2G enabled the replacement of analogy technology with digital communication; while in 3G emphasis is given to data communication, besides the voice communication. Thus, a converged network for both voice and data communication is emerging. Moreover, research is in full swing to introduce many killer applications for the 4G.

The term mobile commerce or M-commerce has no particular definition. Several analysts and vendors have defined M-commerce in varied ways. Put simply, when a mobile device is used for electronic commerce, it is referred to as mobile commerce or M-commerce.

The following are some popular definitions of M-commerce.

(i) M-commerce is the use of a mobile device to communicate, inform, transact, and entertain using text and data via a connection to the public and private network. —*Lehman Brothers*

(ii) Business-to-consumer transaction conducted from a mobile device.

—*J.P. Morgan*

(iii) Mobile commerce refers to any transaction with monetary value that is conducted via a mobile telecommunication network.

—*Duelacher*

Any electronic transaction or information interaction which is conducted using a mobile device and mobile networks (wireless or switched public networks), that leads to the transfer of real or perceived value in exchange for information, service or goods.

—*mobileinfo.com*

The sale of mobile phone ring tones, games, audio and video content, including full-length music tracks are the primary usages of M-commerce. M-commerce also caters to information transfer, such as cricket scores via the short messaging service. The modes of payment used in M-commerce are:

- Deducting a mobile user's their calling credit, either directly or via the reverse-charged SMS
- Premium-rate calling numbers

NOTES

M-commerce is sometimes confused as an extension of e-commerce. M-commerce is regarded as a business opportunity with functions and characteristics that are unique and are not just an extension of an organization's Internet-based e-commerce channel. There are many resemblances between M-commerce and e-commerce. The similarities range from the ability to buy a product or service virtually versus a real environment. Table 14.3 shows the dissimilarities between M-commerce and e-business.

Table 14.3 *M-commerce vs E-commerce*

Technology	E-Commerce	M-Commerce
Device	Personal computers	Smart phones, Pagers, PDAs
Operating System	Windows, Linux, Unix	Symbian (EPOC), Palm OS, Pocket PC
Presentation Standard	HTML	HTML. WML, i-mode
Browser	Microsoft Internet Explorer, Netscape Navigator	Nokia Browser, MS Mobile Explorer and other micro browsers
Bearer Network	TCP/IP and Fixed wire line Internet	GSM, GSM/GPRS, TDMA, CDMA

Wireless Application Protocol (WAP)

The WAP forum has developed an open global specification called the Wireless Application Protocol (WAP). WAP allows the mobile user with a WAP device to access and communicate information and services easily and immediately.

A set of communication protocols specified by WAP standardizes the manner in which radio transceivers, mobile phones and other wireless devices can be used to access the Internet, including e-mail, Internet Relay Chat (IRC) and the World Wide Web (WWW). The traditional desktop Internet is very similar to the WAP model.

Phone.com describes WAP as:

WAP bridges the gap between the mobile world and the Internet as well as the corporate intranets and offers the ability to deliver an unlimited range of mobile value-added services to the subscribers—independent of their network, bearer and terminal. Mobile subscribers can access the same wealth of information from a pocket-sized device as they can from the desktop.

Mobile devices must have the following features:

1. Less Memory
2. Less Powerful Central Processing Units
3. Smaller Display
4. Limited Input Facilities
5. Restricted Power Consumption

NOTES

A wireless data network must have the following features:

1. More Latency
2. Less Bandwidth
3. Less Predictable Availability
4. Less Connection Stability

WAP is an attempt to design a protocol keeping in mind the above factors relating to mobile devices and wireless data networks.

WAP is a set of protocols that specifies a complete framework for mobile Internet access. The WAP forum is an industry association founded in 1997 by Nokia, Motorola, Ericsson and Phone.com. The purpose of WAP is to expand the usages of wireless data by providing a platform for developing new value-added services. WAP brings superior services and content to digital mobile phones and other mobile terminals from the Internet.

The Basic WAP Model

WAP has three main components that are necessary for the protocol to function properly.

WAP Components

1. WAP Gateway
2. WAP Server
3. Client Device

1. WAP Gateway

The WAP gateway provides communication between wireless networks and the Internet. It acts as an interface between telecommunication protocols within the mobile operator's network and Internet protocols.

The main benefits of a WAP gateway include:

(i) The WAP gateway translates the protocols between the Internet and the wireless network.

(ii) WAP content is encoded into a compact binary form by the WAP gateway over air transmission due to efficiency reasons.

2. WAP Server

The second important component is the WAP server. It stores the WAP content. The WAP server can be on a local network or anywhere on the Internet. WAP servers contain Wireless Markup Language (WML), WML Script and Wireless Bitmaps which are all WAP content formats.

3. Client Device

The client device can be a mobile phone, a pager, a Personal Digital Assistant (PDA) or any other WAP enabled device.

NOTES

The World Wide Web Model

The Web or the WWW model is used on the Internet to provide the client with the ability to receive content in a defined data format from Web servers. The standard protocols, like the Transmission Control Protocol and the Internet Protocol (TCP/IP) and Hypertext Transfer Protocol (HTTP) handles the communication. Figure 14.6 shows that to reach the content on the server a Uniform Resource Locator (URL) is used by the clients.

The client uses a Web browser to display the contents provided by the server in a standard format called the Hyper Text Markup Language (HTML) and the script language (Java Script, Visual Basic (VB) Script, etc.,) to enhance the content functionality.

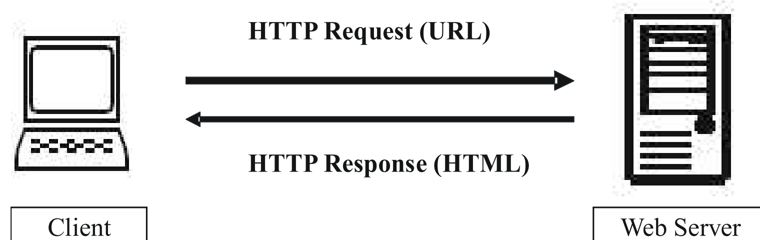


Fig. 14.6 *World Wide Web Model*

The WAP Model

WAP is designed to use existing technology and standards. A browser in a WAP device communicates with the WAP gateway connected to the Internet. The WAP gateway translates requests from the www protocol to stack the WAP protocol and vice versa.

When the browser sends a request to the server, WAP gateway decodes it to plain text and sends the request to the WAP server as the desired content, as show in Figure 14.7. By this method, a content provider only needs to add a few content types to the WAP server to enable the WAP services to be developed, as the user of the WAP device is always connected to same gateway. The Wireless Markup Language (WML) script is the standard content format used by WAP applications. When a server replies, the desired content is sent to the gateway. The content is encoded by the WAP gateway into binary form of WML and sent to the WAP device. This binary encoding is used to reduce network traffic. The textual content is not only compressed, but also all unnecessary spaces and line breaks are removed in this way.

NOTES

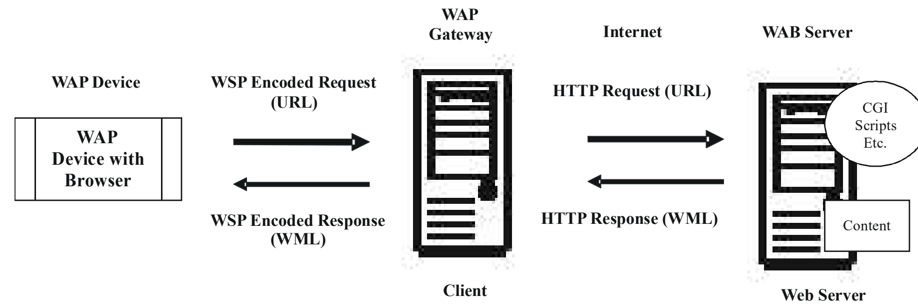
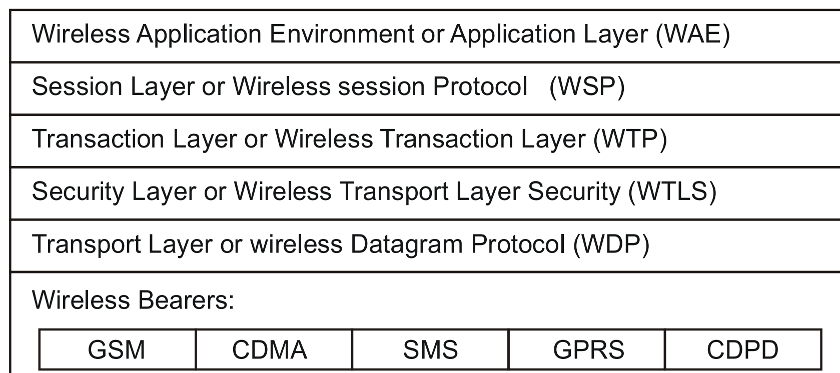


Fig. 14.7 *The WAP Model*

The WAP client communicates with the WAP server by using the WAP gateway. The WAP gateway translates the client requests to www requests before submitting the request to the WAP server.

The WAP Protocol Stack



Wireless Application Environment or Application Layer (WAE)				
Session Layer or Wireless session Protocol (WSP)				
Transaction Layer or Wireless Transaction Layer (WTP)				
Security Layer or Wireless Transport Layer Security (WTLS)				
Transport Layer or wireless Datagram Protocol (WDP)				
Wireless Bearers:				
GSM	CDMA	SMS	GPRS	CDPD

Fig. 14.8 *The WAP Protocol Stack*

Figure 14.8 depicts the wireless application protocol in a series of layers. The Open Systems Interconnection (OSI) model defines a layered framework for generically describing and designing protocols. Similarly, WAP has six layers. Each layer is responsible for managing some part of WAP. The WAP architecture layers are analysed as follows:

1. Application Layer or Wireless Application Environment (WAE)

In the WAP protocol stack the WAE is the uppermost layer. WAE is divided in two logical layers:

(i) The First Layer is for User Agents

(ii) The Second Layer is for Service and Formats

The first layer is for user agents, that is browsers, phone books and message editors. In the second layer, there are different elements such as WML, WML Script, image formats, card and calendar format, and so on.

NOTES

WAE provides an important service, that is the URL service. The URL identifies resources on a server that can be reached by well-known protocols. WAE also supports Uniform Resource Identifier (URI), which is used to locate resources that are accessed without using well-known protocols, such as wireless telephony functions.

2. Session Layer or Wireless Session Protocol (WSP)

Two interfaces for the WAE are provided by the wireless session layer.

(i) A connection-oriented service that functions on top of the transaction layer protocol.

A session is provided by the connection-oriented service between the WAP gateway and the client. It takes care of communication interruptions, like change of bearer and capability negotiation.

(ii) A service that is connectionless operates above a non-secure or secure datagram service. The WAE uses a thin layer when a reliable transaction of data is not required.

In short, a session layer provides the following services:

(i)) It creates and releases a connection between the client and the server.

(ii) Exchanging data between the client and the server using a coding scheme that is much more compact than the traditional HTML text.

(iii) Suspending and releasing sessions between the client and the server.

3. Transaction Layer or Wireless Transaction Protocol

A lightweight transaction-oriented protocol is provided by WTP which is appropriate for mobile phones and which also runs on top of a datagram service. A dependable way of communication is provided with an ability to avoid duplication and retransmit lost messages. WTP manages three types of transactions:

(i) **Unreliable One-Way Request:** A message is sent and no acknowledgement is expected from the receiving device. Such applications use the Wireless Datagram Protocol (WDP).

(ii) **Reliable One-Way Requests:** A message is sent and the recipient sends an acknowledgement.

(iii) **Reliable Two-Way Request-Reply Transactions:** A message is sent and the recipient replies with exactly one result message. The sender then finally acknowledges the result message. If the recipient knows that the message processing time exceeds the initiator's timer interval, the recipient may send a 'Hold On' message to prevent the initiator from resending the original message.

4. Security Layer or Wireless Transport Layer Security (WTLS)

NOTES

A transport layer security between the WAP gateway and the WAP client is provided by the security layer. Your data is protected by WTLS including privacy, denial-of-service protection, authentication and data integrity. Data integrity ensures that it is unmodified and uncorrupted. WAP privacy services guarantee that all transactions between the WAP device and the gateways are encrypted. Authentication and protection against denial-of-service attacks are also parts of WTLS.

5. Transport Layer or Wireless Datagram Protocol (WDP)

WDP provides an interface that is consistent to the stack's upper layer and the datagram layer forms the base of the WAP protocol stack. The WDP layer is not required if WAP is applied over a bearer supporting User Datagram Protocol (UDP). WDP also allows correction of data error. On GSM SMS and other bearers, provided by WDP provides the datagram functionality. If needed WDP can extend with the functionality for re-assembling and segmenting datagrams that are large for the underlying bearer.

6. Wireless Bearer Network

The wireless bearer networks are at WAP's lowest level. To function on various bearer services the WAP protocols are designed for them, such as packet-switched networks, short message services and circuit switched connections. Each network has its own advantage and disadvantage in terms of performances, delays and errors.

Generations of Mobile Wireless Technology

First Generation (1G)

- The First Generation (1G) of mobile phones can be traced to the early 1980s.
- The first generation of wireless technology used the analog technology.
- It used FDMA technology for modulation.

Advanced Mobile Phone Service (AMPS) was used in the US and Australia. Nordic Mobile Telephone (NMT) was used in Switzerland, the Netherlands, Europe and Russia.

Problems with First Generation

The problems in the first generation are as follows:

1. The problem with AMPS is that in the 800 MHz band, radio waves are 40 cm long and travel in straight lines.
2. Signals are absorbed by trees and plants and high buildings.
3. It has creates a high level of echo and signal distortion.

4. It has a low calling capacity.
5. It allows minimum privacy.
6. It provides poor data communication capability.

Second Generation (2G)

- Second Generation or 2G technologies use digitized technology.
- It uses a combination of TDMA (Time Division Multiple Access) and FDMA (Frequency Division Multiple Access) technologies.
- Four 2G systems are in use and are:

D-AMPS	–	Digital Advanced Mobile Phone Service
GSM	–	Global System for Mobile
CDMA	–	Code Division Multiple Access
PDC	–	Personal Digital Cellular

Among these, GSM and CDMA are the popular systems.

In 2G technology, voice is digitized over a circuit in 1G and 2G networks and data is transacted over a circuit. This technology is called Circuit Switched Data (CSD). Using modems, a data connection is established between the device and the network.

2.5 Generation (2.5G)

- In the 2.5 technology, voice is digitized over a circuit. However, data in 2.5 G is packetized.
- 2.5G uses the same encoding techniques as 2G.
- General Packet Radio Service (GPRS) networks is an example of 2.5G.

Third Generation (3G)

- 3G technology uses spread spectrum techniques for media access and encoding.
- In a 3G network, both data and voice use packets.
- The International Telecommunication Union (ITU), as a part of its International Mobile Telecommunication 2000 (IMT-2000) programme, defined the technical framework of 3G.
- The two 3G models of CDMA are wideband CDMA (WCDMA) and CDMA 2000.

Components of Mobile Commerce

There are two main components of mobile commerce. These are:

1. Wireless communication component
2. Access device for mobile computing

1. Wireless Communication Component

NOTES

Wireless communication component is a medium through which signals cannot be guided through a solid medium. In a wireless system, air is the medium through which electromagnet solid energy can flow easily in several directions via antennas.

The wireless (unguided media) methods include radio-based, satellite-based and light-based wireless systems.

- **Radio-Based Wireless System:** Radio-based systems include radio waves, satellites and paging systems. These use radio frequency signals that range between 100 KHz – 20 KHz.

Radio waves include the following types:

(i) Short Wave

(ii) Ultra High Frequency

Radio Frequency (RF) is of three types:

(i) One-way RF signals that cover upto 10 miles; for example, digital pagers.

(ii) Two-way Transceivers are used by RF for sending and receiving data upto six miles.

(iii) Local: It functions in the Medical, Scientific and Industrial (ISM) bands.

Radio-based services can be grouped into two main categories:

(a) Land based: Cellular communication packet data between network and Specialized Mobile Radio (SMR).

(b) Satellite based: Paging systems and Very Small Aperture Satellite (VSAT).

Land-based radio systems use low power transmitters and receivers to send and receive data, while satellite-based systems use high frequency.

- **Satellite-Based Wireless System:** Satellite microwave systems transmit signals between directional parabolic antennas. A communication satellite is a microwave relay station placed 36,000 km above the equator. It is a satellite placed in a geosynchronous orbit. It appears to be stationary relative to the earth and always stays over the same point with respect to the earth.
  - o Satellite frequency range lies between 4–6 GHz and 11-14 GHz .
  - o Satellite microwave installation for orbiting satellite is technical and difficult.
  - o Satellite network is useful in any situation where data needs to be dispersed.
- **Light-Based Communication System:** Light-based communication system use light to send pulses that emit in a Light Emitting Diode (LED) to a photo sensor that decodes the signals.

A light-based communication system has two categories:

(i) Low-Speed Devices (Range 115 kbps to 250 kbps)

(ii) High-Speed Devices (Range 1.2 mbps)

- The installation cost of a light-based system is less and the process is hassle free. It is also capable of sending data multiple times. It is still not as fast as many cable connections, but it is fine for updating files, downloading the day's schedule or transmitting other forms of data. The goal is compatibility between personal digital assistants, notebooks and desktop computers as well as peripherals, such as printers, fax machines – and even telephones.

NOTES

2. Mobile Information Access Devices

There is a wide variety of information access devices. These are:

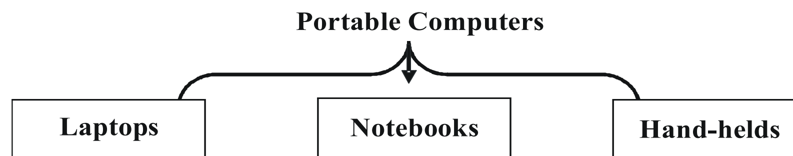
(i) Portable Computer

(ii) Hybrid Computers

(iii) Personal Digital Assistants

(i) Portable Computers

Portable computer can be of any of the following three categories.



(a) Laptops

Laptops are useful for running applications that demand very powerful hardware, such as Computer Aided Design (CAD) and video presentations.

(b) Notebooks

Notebooks serve the traditional user who wants to do a large amount of word processing and manipulate sizable spreadsheets. The subnotebook allows a close approximation of the desktop computing experience and is a mainstream of mobile computing. Subnotebook designs include applications in ROM and Windows Software.

(c) Hand-Held System

Hand-held system is a powerful system when several of its capabilities are combined with more intuitive functions, such as phone lists and messaging.

(ii) Hybrid Pen Computer

A hybrid pen computer provides a pen-based system instead of a keyboard. It is provided with two options:

NOTES

(a) Without Keyboard

(b) With Keyboard

In case a keyboard is provided, it is of the standard size and there is a mouse which is used as a pen. The user gets a tablet-size writing pad when the display is closed.

(iii) Personal Digital Assistants (PDAs)

Personal digital assistants are similar to PCs but smaller in size can fit inside a coat pocket. The PDA is optimized in size and weight, which means it has limited functionality. The PDA market is divided into three functional segments:

(a) **Digital Assistants:** In this segment, the hand-held device captures data and digitizes it. It combines personal information management with wireless voice, data and fax communication. They rely on a pen-based user interface in place of a keyboard and makes heavy use of handwriting recognition. Digital assistants include a wide array of information management tools and notepad and sketchpad.

(b) **Personal Communication:** In this segment, the mobile telephone has an LCD screen integrated into it. Personal communicators are suitable for users who want two-way communication and an all-in-one device. But when an LCD screen is attached to the mobile phone then it is difficult because a user cannot read the screen and talk on the phone at the same time. It is expected that in some time, the personal communicator and the digital assistants will come together and will be obtainable in various forms.

(c) **Palmtops:** In this segment, there is a large storage segment (probably on CD-ROM) for the user who wants to carry a lot of data and is also able to retrieve it easily. Palmtops give support to tasks that are personal, like simple note taking, diary maintenance, expense tracking and many types of calculations. Communications terminal allows accessibility to e-mails and palmtops provide collective functionality of a personal organizer.

Networking Standards for Mobiles

There are three basic types of mobile networking standards:

1. Time Division Multiple Access (TDMA)
2. Frequency Division Multiple Access (FDMA)
3. Code Division Multiple Access (CDMA)

1. Time Division Multiple Access (TDMA)

In TDMA, the radio channel is divided in a way that users are able to transfer information in different time slots, to enable multiple users to share the channel. TDMA is the first digital system standardized in North America. In TDMA, multiple users, data services or sources are allotted different time-slices to access the same channel. The available time slice is divided among multiple modulated-signal sources. Therefore, each transmitter uses the channel after the time interval of (577 × 8) microseconds or 4.615 ms (GSM).

2. Frequency Division Multiple Access (FDMA)

Frequency multiplexing requires a separation of the frequency bands used for transmission by different channels. The available frequency range is divided into bands which are used by multiple sources or channels at the same time. FDMA is an access method which assigns different slices to different users for accessing the same carrier. FDMA suffers from inter-modulation.

3. Code Division Multiple Access (CDMA)

In CDMA neither the frequency domain nor the time domain are divided, rather the same frequency is shared at the same time by all users. A code or sequence is assigned to every user and the bit rate of the information transferred by the user is bigger than the bit rate of the sequence.

As long the correct code is with the receiving device, it can select its conversion out from others. Any further user that is added, the complete interface enhances and it gets difficult to take out a user's sequence that is unique to him from other users. For CDMA, maximizing the battery life of mobiles and the precise power control is significant in increasing the system's capacity. Mobiles that transfer extreme power boost interference to other mobile phones.

Future Trends in Mobile Commerce

Most trends and advances are seen in Asia, Europe and North America. But nowadays, almost all countries benefit from using M-commerce. Many businesses use M-commerce as a competent process of meeting the demand of their customers. That is the reason why mobile phones and PDAs have become so popular.

M-commerce has been adopted by banks and many financial institutions to access account information, for example, stock quotes, financial advices and balance inquiries from anywhere and make transactions, like paying money and buying stocks through the mobile phone. This method is termed as m-banking or mobile banking.

Mobile phones are also becoming popular for providing news from anywhere in the world. The demand for mobile services has led to the growth of sports, entertainment, reservation and shopping service on the mobile phone.

NOTES

M-commerce is being used by organizations to increase the scale of everything and widening the range—from advertising, marketing and services. Security plays an important role in mobile commerce. Mobile companies, today, invest a significant part of their revenue for the protection of their customers and information related to them from intruders and online hackers.

Advantages of M-Commerce

The advantages of M-commerce are:

1. The mobile phone can be personalized as per the mobile owners' preference and also have full control over data.
2. Cost savings, innovative business opportunities and customer satisfaction can be ensured.
3. Improved, quicker and efficient linkage with clients can be provided by M-commerce.
4. The buyer and the seller can be brought closer through M-commerce thus facilitating higher profits and better customer relationship.
5. With the help of the light-weighted device, M-commerce can be used anywhere and anytime.

Disadvantages of M-Commerce

The disadvantages of M-commerce are:

1. Since mobile phones come with small screens; it limits the intricacies of the applications.
2. Mobile phones contain graphics or processing power that are limited in comparison to a PC.
3. Some overheads are added due to security problems.
4. Different networks have different approaches towards M-commerce.

14.7 E-COMMERCE APPLICATIONS IN INDIA

E-commerce (Electronic commerce) is the activity of electronically buying or selling of products on online services or over the Internet. Electronic commerce draws on technologies, such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, OnLine Transaction Processing (OLTP), Electronic Data Interchange (EDI), inventory management systems, and automated data collection systems. E-commerce is in turn driven by the technological advances of the semiconductor industry, and is the largest sector of the electronics industry.

India has big number of Internet users.

NOTES

Challenges of Traditional Marketing

Traditional marketing operates on the basis of the four Ps, that is, the correct marketing mix of product, price, promotion and placement. Before launching any product, its production and management, price, promotion, placement or distribution, its retailing and the procedure by which it is delivered to the end-user all need to be planned, decided and taken care of. The target customers need to be segmented and studied so that a detailed marketing plan can be drawn up. Then, an advertising plan is decided upon.

The challenges faced by traditional marketing are as follows:

1. **Expensive Product:** When a company decides the cost of a product, it includes all expenses like product information brochure, shipping charges, mailing and service of human resources, among others. As a result, the cost of product is high.
2. **Lack of Consumer Interaction:** It is not possible for the company to interact with the consumer because between the two, operate the wholesaler and retailer.
3. **Lack of Coordination:** Sometimes, companies face some problems in marketing activities because marketing executives, brochure printers and advertising agencies, etc, lack coordination and all of them cannot perform at the same time because all work has to be passed by management, channel wise. Therefore, delays are inevitable.

With customers being exposed to more and more brands with each passing day, brand recall is reducing. It is difficult for consumers to make a brand choice as all brands are as good as their competitors. This leads to a need for an all-round marketing strategy which covers all the existing modes and also aims at constant brand recall. This is where the reach of the Web plays a significant role.

Retailing in E-Business Space

Doing business in cyber space is advantageous to retailers in many ways. It reduces labour costs and helps avoid expenditure on paperwork and merchandising. It is a known fact that online transaction costs are lower. Some of the other merits are as follows:

Better Customer Service and Satisfaction

Customer relationship management on the Net is more efficient and gathering of information is a lot more easier. The collected information can be used for customer service. It is easy to gain background information related to the location, recent buys, account history and status of payment.

It is possible for customers to track their own orders, search for the nearest outlet or find answers to their queries. Customer information and data related to purchasing are used to guide marketing and purchasing and to improve the quality of services.

NOTES

Personalization of Services

Services are personalized and marketing is optimized. Several different communication channels including live chat, e-mail, telephone and online FAQs are combined to assist customer relationship management. Customers have the option to select the method in which they want to do business with you. You will be aware of the background of your customer, the time at which he placed the order, the nature of his complaint, irrespective of when he/she made it.

Reduction in Customer Service Cost

As customers are able to search for the products they want, on their own, they are able to get exactly what they are looking for. They even check their order status on their own. The emphasis on self-service helps reduce the expenditure on customer service. Websites have customized display of product range with the prices and relevant information for the selected product categories. The content is usually available in multiple languages for the convenience of customers.

Benefits of an Online Catalogue

Online catalogues are regularly updated with new models, products, pricing descriptions or offers. Printing and mailing costs can be avoided and you can indulge in short-term price promotions. Online catalogues allow interested customers to look at the products within their specific price range. Link to inventory allows customers to confirm the availability of products and the time of delivery. Special customers can be offered discounts or rewards on the basis of their online purchase history or current interest through special promotions, free delivery offer (over a certain sales limit), coupons and frequent-shopper points.

Internet Marketing

The term Internet marketing is interchangeably used for Web marketing, e-marketing, Internet advertising or online marketing. The wide availability of the World Wide Web allows businessmen across the world access to millions of potential customers. All that is required is a few clicks of the mouse. Thanks to the Net, your products and services can be viewed or read about by many people in different parts of the world, at the same time.

Services of professional Web marketers are available these days who ensure that potential customers not only visit your site, but also become regular customers.

Once you subscribe to such services, they help to market your product through various facilities such as the Mall Linking Service. This enables the client's

site to get linked to the Internet connection of numerous other online shopping mall service providers. They log on to the sites provided and managed by these marketing specialists. They promote the website according to the nature and function of search engine movements. Marketing techniques such as Advanced Press release services, Click Exposure techniques and Advanced Search Engine facilities are used. Click exposure techniques function through direct targeted keyword advertising. Online marketing services target a chosen category for advertising using effective pop-under. This is qualitative, targeted and scaleable website traffic which can help your business to grow very fast.

NOTES

In the world of advertising, Internet marketing is one of the popular sources for promoting business, business products and services because of its ability to reach the target audience. An online shopping website is of course an advantage as the facility will then be open all year round and for twenty-four hours a day. Professional Internet marketers will ensure that your business or firm appears right on the top of important search engines and directories. A part of the role of a professional Internet marketer is to make the website design user-friendly, optimize the search engine, optimize the structure of the website, optimize keyword, etc.

Types of Internet Marketing

There are many types of Internet marketing. Some of these are discussed as follows:

1. Pay Per Click
2. Search Engine Optimization (SEO)
3. Internet Auctions
4. Affiliate Marketing
5. Banner Advertising
6. Directory Listings
7. Ethical E-mail Marketing
8. Viral Marketing

Let us learn a little about each of these types.

- **Pay Per Click:** In this type of marketing, a certain amount is paid to the search engine by the website, to ensure that the site is featured separately whenever keywords related to that particular sector or industry are typed in.
- **Search Engine Optimization (SEO):** This implies the making of a website more search engine-friendly. As a result, whenever a user types a search or a specific keyword, the resultant site containing the keyword or search item will show in the top search results. To attract potential customers, this kind of Internet marketing is perfectly suitable.

NOTES

- **Internet Auctions:** Online auctions are similar to the auctions that most of us have seen or are familiar with. Items are put up to be sold at a minimum price quoted by the seller. Those interested in buying start quoting the price they are willing to pay for them. The person who quotes the highest price is known as the highest bidder, and he gets the item. A buyer can be from any part of the world in the case of online auctions.
- **Affiliate Marketing:** When a firm engages other websites or firms to market their products, it is termed as affiliate marketing. These hired firms or websites are referred to as affiliates. It is the responsibility of the affiliates to market the products of the firm that hired them. These affiliates may carry or display the company's logos or banners on websites hosted by them. Visitors who click on these banners are automatically redirected to the sponsor company's site. For this work, the affiliate receives a commission.
- **Banner Advertising:** This kind of advertising involves the displaying of banners of the advertisers on the website. Displaying the banner on the website for a particular period of time involves a certain amount of money to be paid by the advertiser.

  Banners are designed to be visually attractive in order to get the customers' attention. The website on which the banner has been displayed decides the kind of response it gets. However, it cannot be guaranteed that the banner will be viewed by the prospective customers.
- **Directory Listings:** When websites are placed in particular categories in a directory, this service is termed as directory listings. This service can be chargeable or provided free of cost. Yellow Pages and Yahoo are examples of directories where such listings can be done. Therefore, Internet marketing is a new age marketing which allows sellers with good prospects. Internet marketing is cheaper than traditional marketing and allows even small-scale businesses to thrive and flowish.
- **Ethical E-Mail Marketing:** Several Web applications allow website owners to maintain the databases of subscribers and also provide the facility of sending personalized newsletters within no time. Usually, marketing strategies are impersonal. E-mail marketing provides a chance to address the recipient personally. If e-mail marketing is handled tactfully, it can be utilized to appeal the target audiences' emotions and obtain an optimistic reaction from them. Information, about new products and launches, seminars, up-to-date information on service workshops, etc., can be conveyed to the subscribers with the help of e-mails
- **Online News Articles:** Online news articles complement SEO and online public relations. News helps spread consumer awareness and converts the

reader into a prospective buyer. This will result in an increase in the brand value of e-business. Professional agencies help their clients/website owners to prepare news articles and upload on relevant sites and online news media.

- **Viral Marketing:** Viral marketing is when a customer buys a certain product and is convinced to buy something else too in the process. When you visit a shop, for instance, to purchase a bed, the salesman will try to sell you matching sheets, cushions, pillows, mattresses, etc. Similarly, on the Net, when people send e-mails, they unwittingly end up infecting the recipient of their mail with the ads appearing at the bottom of the message. This is called viral marketing.

Internet Consumer and Market Research—Needs of Website Consumers

While designing a website, a company may have many objectives in mind. An effective website would attract consumers and describe the objectives of the company such as:

1. Creating a healthy relationship with consumers
2. Attracting consumers to the website
3. Designing an interesting website to encourage consumers to explore
4. Motivating consumers to stay on the website
5. Advising consumers to get more information by clicking on the website links

Needs of Website Consumers

When a company's website is designed, then it should:

1. Provide full knowledge of the products or services offered
2. Facilitate easy purchase of the products or services offered
3. Declare full information of product or service warranties, guarantees
4. Offer a complete company profile
5. Offer information related to the person responsible

There are professional agencies who help website owners examine Web audiences by behaviour, size and demographic profile. Their custom research and analyses can not only help a website owner measure the audience, but also evaluate the site, track the traffic on competition's sites, evaluate the strategies of competitors and accordingly update and develop high-quality/effective content.

Research on consumers generally involves the following:

- Identification of the audience
- Segmentation of visitors to the website

NOTES

- Study of the level of Internet usage by visitors, whether it is light, medium or heavy
- Integration of data pertaining to consumer profiles, sales and attitudes
- Comparison of online and offline performances
- Study of the channels that are used frequently
- Exploration of the ways in which offline and online channels can be used together to gain an edge and increase sales
- Use of custom surveys
- Evaluation of customer satisfaction levels
- Identification of the most effective ways in which customers can be attracted

Advantages and Disadvantages of Internet Marketing

Internet marketing has its positive as well as negative aspects. Both the merits as well as the demerits have to be understood well before a proper Internet marketing strategy can be created.

Advantages

- The online store is open at all times, day and night. Customers from all parts of the globe can shop at any time of the day or night.
- It is the most economical way of spreading your message. Sending e-mails to subscribers is a lot cheaper than posting letters to them.
- Subscribers can be updated promptly through e-mails. Website visitors can access current and updated information every time they visit the website. Customers can be informed about promotional schemes and sales as soon as they access their e-mails.
- Online magazines/newspapers or law firms are information-sensitive businesses. They can obtain products straight without relying on the courier service. Products can be delivered directly to the customers without using a courier service.

Disadvantages

- Online marketing does not come without a price. The cost of offering the product or providing the service should take into account the expenditure on software, hardware, time, effort, website design and maintenance and online distribution costs.
- Only about 50 per cent of households indulge in online shopping. This means that you are able to reach out only three out of six households.

NOTES

- The Internet is still regarded as a source of information gathering by most customers. Of the total number of browsers who visit websites, the vast majority motivated to purchase will probably buy in person because people are more comfortable with live interaction during sales transactions. Customers do not show interest in purchasing from small businesses that have only one location.
- It is very important to keep updating your site since outdated information can result in losses.
- The traditional form of old-fashioned customer service is still popular and preferred. The majority of marketers online do not have very strong customer service or inquiry response programmes. As a result, many online visitors may rate a website as having poor customer service even before actually coming in contact with the product/service provider.
- Poor navigation facility also makes it difficult for visitors to search for what they want. This happens more often on websites that are designed from the marketing point of view and not from the customer service point of view.
- The security of various sites is doubted by the visitors. Many visitors do not want to risk transacting online using credit cards for fear of having their personal/credit card information stolen.
- Those visiting your site have probably been to other similar sites as well. A purchase can be guaranteed only if they find exactly what they want.
- Visitors expect some free content or special offers or bonuses.

Advertisement and Display on the Internet

Advertising plays a very important role in the overall marketing strategy. It involves bringing a product or service to the attention of potential and current consumers. It involves the process of developing and placing ads, direct mails, billboards, television, radio, Internet, etc. Advertising is usually the largest expense of a marketing plan. Therefore, advertising on the Web goes much beyond simply putting up banners. You have to sell your ad space, ads, put them up and also employ good advertising display and tracking software. The Internet has become the accepted medium of advertising and more businesses are turning towards the Web instead of the traditional newspaper, radio and TV advertising which are becoming unaffordable, especially for small entrepreneurs.

According to the Interactive Advertising Bureau (IAB), and PricewaterhouseCoopers (PwC) that released the IAB Internet Advertising Revenue Report, advertising revenues from Jan to June 2007 were nearly $10 billion. This is a 27 per cent increase than the previous year, at the same time. The advertisers were from consumer advertising sector targeting the increasingly growing online traffic and popularity of several online companies.

NOTES

It is generally known that apart from e-mail being the most used online method, 'Search' is the second. The advertising formats using 'Search', with Google still leads with nearly 64 per cent of the market share mostly in the US, followed by Yahoo. Search ads in Google accounts for 41 per cent of the money spent. It is common knowledge that one of the best ways to get traffic on a web site is through the search engines. This can be accomplished by using blogs effectively. All that is required is indexing of blogs and posting of keywords.

Many types of business benefit from the features offerred by e-advertising. These include consumer businesses, industrial businesses, non-profit organizations, service firms, universities associations and trade groups. Advertising on the Web is a powerful tool for acquiring new customers. Advertising on the Net is developing into a broader mix of factors that an organization employs for enhancing sales even if the business is executed totally offline or fully or partially online. Advertising on the Net is becoming a significant part of almost every organization's marketing mix. When you advertise on the Web, it takes aspects of standard marketing and adds tools like search engine optimization (SEO), advertising, meta tags, and promotion to attracts customers to your website.

E-Advertisement

Another type of direct marketing is e-advertising. This type of marketing makes use of e-mail as a mode of communicating fund-raising or commercial messages to the customers. It must be noted that every e-mail that an existing or a potential customer receives can be regarded as e-advertising.

The Web lowers the cost of publishing and enables even small organizations to become e-publishers.

In e-advertising, the advertising broker accepts advertisements and then places them with suitable websites. The broker gets a payment from the advertiser which he shares with the website publisher. This process can be automated with the advertisers providing a description of the type of websites they would like to advertise on; and this information then being matched with appropriate sites by using meta data related to the geographic region and appropriate keywords.

Auctioning of the websites helps to determine the price of the advertisement. The price is determined by the most popular keywords. This leads to the meta data (about the web pages) and the advertisements becoming commodities and proxies for the ads and web page content. This process is used by the search engine, Google.

So, nowadays when people discuss e-advertising, they generally refer to it as 'online advertising'. This is advertising in the online environment, that is, using websites, e-mail ads, etc. Although the forms may be different, the basic principles remain more or less similar to that of traditional advertising; for instance, organizations make use of paid space to promote businesses.

NOTES

E-advertising is considered as persuasive, non-personal information related to a product or service that appears in an e-mail or on a web site. Businesses have realized that by advertising online, they will be able to reach their target market faster and in a more efficient manner. Online marketers and advertisers have highly benefitted from e-advertising because of cost-effective, innovative and better targeted opportunities.

E-Business for Service Industry

It is not just products that sell well on the Internet, services also benefit from doing electronic business. The hospitality industry, for instance, has been following the trend of online application in recent years. In fact, all travel-related business deals on the Net are referred to as e-travelling.

It is the smaller travel agencies that benefit more from doing business online than the bigger and well-established ones. E-business gives the smaller firms a competitive advantage over their larger counterparts.

Other than hoteliers, travel-related service providers in the travel industry have been quick to take advantage of direct online distribution. All the significant airlines and car-rental firms have adopted the Internet as their main medium of distribution. Most major airlines earn a good part of their revenue from their website dealings. More and more airlines sell directly to online customers and compete with online intermediaries to gain a strong presence on the Internet. Direct Internet consumer sales allows the airlines to cut down on the commission that they would have had to pay to the travel agencies.

Similarly, hoteliers have also realized that online distribution helps cut costs, attract rich customers and reduce dependency on traditional channels that are the more expensive. They have realized that by focussing only on traditional distribution channels, they will only be facing lower occupancy rates and heavy expenditure on distribution and operation costs.

Check Your Progress

1. What are the requirements of e-business?
2. Define the e-business applications infrastructure for online communication and the building blocks of e-business.
3. Explain the concept of online shopping.
4. Explain about the Business-To-Business (B2B) model with its types.
5. What are digital signatures?
6. Explain about Mobile Commerce and Wireless Application Protocol (WAP).
7. How Internet is significant for e-commerce and mobile commerce?

14.8 ANSWERS TO CHECK YOUR PROGRESS QUESTIONS

1. The following are the requirements of e-business:
   - Improved customer service
   - Origin of new business opportunities
   - Enhanced speed and accuracy of a product
   - Product cost saving
2. E-business applications are built on the existing infrastructure for online communication, network and connection software which frames the nascent information superhighway.

   Enterprise applications, insights, functions and IT infrastructure form the building blocks of e-business.
3. Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the Internet using a web browser. Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine, which displays the same product's availability and pricing at different e-retailers.

   An online shop evokes the physical analogy of buying products or services at a regular 'Bricks-and-Mortar' retailer or shopping center; the process is called Business-To-Consumer (B2C) online shopping. When an online store is set up to enable businesses to buy from another businesses, the process is called Business-To-Business (B2B) online shopping. A typical online store enables the customer to browse the firm's range of products and services, view photos or images of the products, along with information about the product specifications, features and prices.
4. The Business-To-Business (B2B) model needs two or more business organizations that do business with each other. It entails commercial activity among companies through the Internet as a medium. At present, there are many types of e-businesses. The B2B e-business is of the following types:

   Supplier oriented: In this type of B2B e-business, a supplier establishes the electronic market where a number of customers or buyers transact with suppliers. Generally, it is done by a supplier which has monopoly over products that it supplies.

   Buyer oriented: In this type of B2B electronic commerce, big business organizations with high volume purchase capacity creates an e-business

NOTES

marketplace for purchases and gains by starting a site of their own. The online e-business marketplace is used by buyers for placing requests for quotations and carrying out the entire purchase process.

Intermediary oriented: In this type of B2B e-business, a third party establishes the e-business marketplace and attracts both buyers and sellers to interact with each other.

5. Digital signatures are used for authenticating e-commerce business transactions. The authentications refer to legal, financial and other document-related issues. Digital signatures are just like handwritten signatures which determine authentications.

   A digital certificate serves as an electronic identity card that establishes the user's credentials when business deals are transacted across the web. A digital certificate is defined as a method to electronically verify for authenticity. The digital certificate is just like an identity card, such as a driver's license.

6. The term mobile commerce or M-commerce has no particular definition. Several analysts and vendors have defined M-commerce in varied ways. Put simply, when a mobile device is used for electronic commerce, it is referred to as mobile commerce or M-commerce.

   The WAP forum has developed an open global specification called the Wireless Application Protocol (WAP). WAP allows the mobile user with a WAP device to access and communicate information and services easily and immediately.

   A set of communication protocols specified by WAP standardizes the manner in which radio transceivers, mobile phones and other wireless devices can be used to access the Internet, including e-mail, Internet Relay Chat (IRC) and the World Wide Web (WWW). The traditional desktop Internet is very similar to the WAP model.

7. India has big number of Internet users. E-commerce (Electronic commerce) is the activity of electronically buying or selling of products on online services or over the Internet.

   Electronic commerce draws on technologies, such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, OnLine Transaction Processing (OLTP), Electronic Data Interchange (EDI), inventory management systems, and automated data collection systems. E-commerce is in turn driven by the technological advances of the semiconductor industry, and is the largest sector of the electronics industry.

   Mobile phones are also becoming popular for providing news from anywhere in the world. The demand for mobile services has led to the growth of sports, entertainment, reservation and shopping service on the mobile phone.

M-commerce is being used by organizations to increase the scale of everything and widening the range—from advertising, marketing and services.

NOTES

14.9 SUMMARY

- These days, consumers want better service. Therefore, e-business services offer a means of communication between the consumer and the company. The consumer can even make online complaints to a company.
- Bigger network between consumers and companies can lead to new business opportunities.
- The usage of e-business services reduces human errors and other problems like a duplication of proceedings. This perfection in speed and accuracy, plus easy access to documents and information affect the increase in production.
- E-Banking facility offers remote banking electronically. Electronic banking is also referred to as online banking, cyber banking, home banking or virtual banking. It enables Web users to make online purchases and pay for the same using an online-banking facility. It is cost-effective, simple and available 24 hours.
- E-business applications are built on the existing infrastructure for online communication, network and connection software which frames the nascent information superhighway.
- Enterprise applications, insights, functions and IT infrastructure form the building blocks of e-business.
- Business intelligence would also be required for Web analysis, for managing knowledge and content and for mining data, if required, depending on the nature of e-business.
- Online shopping is a form of electronic commerce which allows consumers to directly buy goods or services from a seller over the Internet using a web browser.
- Consumers find a product of interest by visiting the website of the retailer directly or by searching among alternative vendors using a shopping search engine, which displays the same product's availability and pricing at different e-retailers.
- An online shop evokes the physical analogy of buying products or services at a regular 'Bricks-and-Mortar' retailer or shopping center; the process is called Business-To-Consumer (B2C) online shopping.

NOTES

- When an online store is set up to enable businesses to buy from another businesses, the process is called Business-To-Business (B2B) online shopping.
- A typical online store enables the customer to browse the firm's range of products and services, view photos or images of the products, along with information about the product specifications, features and prices.
- A mall or shopping centre is a large building that is full of many smaller shops and stores. It is different from earlier markets or bazaars because most of the shops are not little booths or stalls in one big open area. Each store has its own space with walls. Most of their entrances face a central walking area inside the building.
- The Business-To-Business (B2B) model needs two or more business organizations that do business with each other. It entails commercial activity among companies through the Internet as a medium. At present, there are many types of e-businesses.
- In the supplier oriented type of B2B e-business, a supplier establishes the electronic market where a number of customers or buyers transact with suppliers. Generally, it is done by a supplier which has monopoly over products that it supplies.
- In the buyer oriented type of B2B electronic commerce, big business organizations with high volume purchase capacity creates an e-business marketplace for purchases and gains by starting a site of their own. The online e-business marketplace is used by buyers for placing requests for quotations and carrying out the entire purchase process.
- In the intermediary oriented type of B2B e-business, a third party establishes the e-business marketplace and attracts both buyers and sellers to interact with each other.
- The business to consumer model clearly concentrates on individual buyers and is thus known as Business-To-Consumer (B2C) model. The B2C model offers consumers the capabilities to browse, select and merchandise online from a wider variety of sellers and at better prices.
- Cryptography is derived from the Greek words kryptos (hidden, secret) and gráphô (I write). It is the practice and study of hiding information. Cryptography is today considered a branch of both mathematics and computer science, and is used extensively in information theory, computer security and engineering.
- The science of writing in a secret code is called crypto. Cryptography is used in applications which require security of data, such as in the case of ATM cards, computer passwords and electronic commerce.

NOTES

- Data Encryption Standard (DES) is an example of secret key cryptography. It was developed by IBM. DES is block cipher-based scheme which encrypts a 64-bit data block using a 56-bit key. The block is transformed in such a way that it involves sixteen iterations.
- Hash function is a formula that converts a message of a given length into a string of digits called a message digest. A mathematical transformation is used by the hash function to encrypt information such that it is irreversible. The encrypted ciphertext message cannot be decrypted back to plaintext.
- Digital signatures are used for authenticating e-commerce business transactions. The authentications refer to legal, financial and other document-related issues. Digital signatures are just like handwritten signatures which determine authentications.
- A digital certificate serves as an electronic identity card that establishes the user's credentials when business deals are transacted across the web. A digital certificate is defined as a method to electronically verify for authenticity. The digital certificate is just like an identity card, such as a driver's license.
- The term mobile commerce or M-commerce has no particular definition. Several analysts and vendors have defined M-commerce in varied ways. Put simply, when a mobile device is used for electronic commerce, it is referred to as mobile commerce or M-commerce.
- The WAP forum has developed an open global specification called the Wireless Application Protocol (WAP). WAP allows the mobile user with a WAP device to access and communicate information and services easily and immediately.
- A set of communication protocols specified by WAP standardizes the manner in which radio transceivers, mobile phones and other wireless devices can be used to access the Internet, including e-mail, Internet Relay Chat (IRC) and the World Wide Web (WWW). The traditional desktop Internet is very similar to the WAP model.
- The Web or the WWW model is used on the Internet to provide the client with the ability to receive content in a defined data format from Web servers. The standard protocols, like the Transmission Control Protocol and the Internet Protocol (TCP/IP) and HyperText Transfer Protocol (HTTP) handles the communication.
- M-commerce has been adopted by banks and many financial institutions to access account information, for example, stock quotes, financial advices and balance inquiries from anywhere and make transactions, like paying money and buying stocks through the mobile phone. This method is termed as

NOTES

- Mobile phones are also becoming popular for providing news from anywhere in the world. The demand for mobile services has led to the growth of sports, entertainment, reservation and shopping service on the mobile phone.
- M-commerce is being used by organizations to increase the scale of everything and widening the range—from advertising, marketing and services.
- India has big number of Internet users. E-commerce (Electronic commerce) is the activity of electronically buying or selling of products on online services or over the Internet.
- Electronic commerce draws on technologies, such as mobile commerce, electronic funds transfer, supply chain management, Internet marketing, OnLine Transaction Processing (OLTP), Electronic Data Interchange (EDI), inventory management systems, and automated data collection systems. E-commerce is in turn driven by the technological advances of the semiconductor industry, and is the largest sector of the electronics industry.

14.10 KEY WORDS

- **E-advertising:** Advertising information on the Web.
- **E-banking:** Enables users to make online purchases and make payments using an online banking facility.
- **Enterprise application software:** Refers to software that facilitates the performance of business functions such as scheduling, accounting, management of customer databases, and so on.
- **E-business:** A business interaction that is performed by using the electronic medium.
- **Business model:** A set of shared common characteristics, behaviour and methods of doing business that enables a firm to generate profits through increasing revenues and reducing cost.
- **Consumer-to-consumer:** Consumers sell directly to other consumers via online classified advertisements and auctions or by selling personal services or expertise online.
- **Cryptography:** The science of writing in a secret code
- **Encryption:** A cryptography technology to scramble (encrypt) the data with a key so that no one can make sense of it while it is being transmitted
- **Intruder:** Any person who does not have the authorization to access the network or the information
- **Plaintext:** An unreadable message that needs to be converted into an intelligible message or encrypted message.

NOTES

- **Ciphertext:** A message in an encrypted form.
- **Hash function:** is a formula that converts a message of a given length into a string of digits called a message digest.
- **Non-repudiation:** is the activity of presenting of tamper-proof evidence proving that a message was sent or received
- **Mobile commerce:** Using a mobile device, such as a mobile phone or PDA to conduct e-commerce.
- **Wireless Application Protocol (WAP):** A set of protocols that specifies the complete framework for mobile Internet access
- **3G technology:** Uses spread spectrum techniques for media access and encoding
- **WAP gateway:** Provides communication between wireless networks and the Internet
- **Satellite-based wireless system:** Satellite microwave systems transmit signals between directional parabolic antennas.
- **Internet marketing:** Also called Web marketing, online marketing or e-marketing, it involves marketing of products on the Internet so that the business is promoted to a larger audience, spread across the globe.
- **Affiliate marketing:** A type of marketing wherein a firm hires other firms or websites for marketing products

14.11 SELF ASSESSMENT QUESTIONS AND EXERCISES

Short-Answer Questions

1. Define the term e-business.
2. What is electronic trading and marketing?
3. Define the significance of online shopping and malls.
4. Explain about the B2B and B2C models.
5. Why document and transaction security is important?
6. What is the importance of digital signature?
7. How integrated transactions is done on mobile platforms?
8. List some e-commerce applications of India.

NOTES

Long-Answer Questions

1. List the various prerequisites of e-business.
2. Explain the various functions of e-business. Also, add a note on the types of shopping service and information services.
3. Formulate a model of the architecture of e-business framework. What are the building blocks of e-business?
4. Discuss the main features of the three fundamental types of business models.
5. Describe the six components of business models.
6. Summarize the role of a business model.
7. Discuss the meaning and purpose of cryptography.
8. Compare the features of secret key cryptography and those of public key cryptography.
9. Describe how digital signatures are validated.
10. Explain the method of ensuring non-Repudiation and message integrity of message.
11. Describe the architecture of the WAP model. Add a note on the significance of each layer of the model.
12. Trace the evolution of the various generations of mobile wireless technology. What are the advantages of one generation over its preceding generation?
13. Identify and list the various advantages and disadvantages of M-commerce.
14. Identify and list the various challenges faced by traditional marketing.
15. Explain the features of the different types of Internet marketing.

14.12 FURTHER READINGS

Rajaraman, V. 2010. *Fundamentals of Computers*, 5th Edition. New Delhi: Prentice Hall India Learning Pvt. Limited.

Sanders, Donald H. 1983. *Computers Today*, 1st Edition. New York: McGraw-Hill.

Krishnamurthy, E. V. and S. K. Sen. 1976. *Computer-Based Numerical Algorithms*. New Delhi: Affiliated East-West Press Pvt. Ltd.

Lipschutz, Martin M. and Seymour Lipschutz. 1981. *Schaum's Outline of Theory and Problems of Data Processing*. New York: McGraw-Hill.

NOTES

Sinha, Pradeep K. 1992. *Computer Fundamentals*. New Delhi: BPB Publications.

Wirth, N. 1988. *Algorithms Plus Data Structures Equals Programs*. New Delhi: Prentice Hall India Learning Pvt. Limited.

Mishra, S. K. and J. C. Binwal. 1991. *Computer Applications in Social Science Research*. New Delhi: Vikas Publishing House Pvt. Ltd.

Gujarati, Damodar N. 1995. *Basic Econometrics*. New York: McGraw-Hill.

Johnston, Jack. 1996. *Econometric Methods*. New Delhi: Tata McGraw-Hill.

Agrawal, Rajneesh and Bharat Bhushan Tiwari. 2008. *Data Communication and Computer Networks*. New Delhi: Vikas Publishing House Pvt. Ltd.